THE ROLE OF SURFACE PLASMON-POLARITONS IN THE PROPAGATION OF LIGHT IN SUB-WAVELENGTH SLITS

by

Jeff Wuenschell

BPhil in Engineering Physics, University of Pittsburgh, 2005

Submitted to the Graduate Faculty of

the School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

SCHOOL OF ENGINEERING

This thesis was presented

by

Jeff Wuenschell

It was defended on

July 5th, 2007

and approved by

Joel Falk, Professor, Department of Electrical and Computer Engineering

William Stanchina, Professor, Department of Electrical and Computer Engineering

Thesis Advisor: Hong Koo Kim, Professor, Department of Electrical and Computer Engineering

THE ROLE OF SURFACE PLASMON-POLARITONS IN THE PROPAGATION OF LIGHT IN SUB-WAVELENGTH SLITS

Jeff Wuenschell, M.S.

University of Pittsburgh, 2007

Over the past few decades optics has begun to take over some of the duties of electronics. The primary focus of recent study in electronics has been miniaturization; meanwhile, a similar trend has begun in optics research with the growth of the nano-optics field. One of the key interests of this field is surface plasmon-polaritons, propagating light which is bound to the interface between a metal and a dielectric.

Electromagnetic waves in the form of surface plasmons break some of the rules of classical optics. Classical optics predicts that light cannot propagate through an aperture much smaller than about half its wavelength – the best it can do is "tunnel" through, which tends to result in weak transmission when the aperture is reasonably long. Experimentally, it has been shown that the wavelength dependence is significantly different than predicted by classical theory. Some structures have been shown to produce even more astonishing results; for example, a thick nano-slit array can exhibit near 100% normalized peak transmission.

The purpose of this thesis will be to analytically and numerically study light propagation in sub-wavelength metallic slits. It will be shown that symmetric surface plasmon modes are the primary carriers of electromagnetic power in a sufficiently small slit. The dynamics of the power and polarization charge inherent in the propagation of these modes will be analyzed. Finally, these topics will be discussed with respect to possible applications, focusing on the utilization of sub-wavelength metallic slits for chemical sensing.

TABLE OF CONTENTS

LIST OF FIGURES

PREFACE

For analytic calculations I will be using the SI version of Maxwell's equations (to be discussed when appropriate). For all FDTD results and explanations, Gaussian units (wherein the units of the electric and magnetic fields are the same) will be used. This is primarily to make it easier to visually compare the fields. Unless otherwise explicitly stated, any epsilon other than ε_0 (such as ε_M and ε_D) will refer to the relative dielectric constant. As a matter of convention, complex plane waves will be expressed as $e^{i(kx-\omega t)}$ (rather than $e^{i(\omega t-kx)}$). While there is no physical difference between the two, some equations may differ slightly from certain references which use the opposite convention (for example, Yariv and Yeh [1]) – the difference in these cases will only be in regards to whether the imaginary part of the result is positive or negative.

I want to thank Joel Falk and William Stanchina, for taking the time to serve on my committee and to provide their comments and criticism of my thesis.

I want to thank my advisor, Hong Koo Kim, for providing me with the right balance of guidance and the freedom to choose the direction of my work. The time I've spent in his group has been an invaluable experience.

I'd like to acknowledge Yun-Suk Jung, whose experimental work ran parallel to my analysis. Many of the otherwise un-cited references to experimental observations in my thesis can be attributed to him. Zhijun Sun – Dr. Kim's former doctoral student - also deserves

recognition; his work was an indispensable reference and guide when I began my own journey into this field.

Most of all, I would like to thank my family – particularly my parents – for their support in this and all endeavors.

This work has been funded by the US National Science Foundation under the grants NIRT-ECS-0403865 and ECS-0424210.

1.0 INTRODUCTION

1.1 PLASMA OSCILLATIONS

By definition, plasma is essentially a gas of unbound, charged particles, usually ionized atoms. This state of matter can exist in many, very different contexts. At sufficiently high temperatures, the average kinetic energy is enough to tear electrons and protons apart and plasma is formed (as in stars). At sufficiently high electric fields, the electrons and protons are likewise torn apart and plasma is formed (as in lightning). Plasmas can occur in solid state systems as well, but for more subtle reasons than in lightning or stars. Metallic atoms have low ionization energy, indicating that the outermost electrons are weakly bound. When a large number of atoms are packed close together – as in a crystal lattice – the valence electrons have no problem hopping from one atom to the next and are free to move throughout the lattice, effectively creating a plasma. The effect of the periodic potential on the electrons is approximately equivalent to just changing the mass of the electrons; in other words, the electrons in the lattice can be approximated as particles in free-space with a different mass.

Electrons in a bulk metal will tend to have an equilibrium density, where the repulsive forces associated with the electron-electron interactions are balanced with the attractive forces associated with the positive background charge of the lattice. In short, the system is essentially charge-neutral. If there is a slight disturbance from this equilibrium, the tendency of the system

will be to restore it. Electrons will want to leave a region with higher than equilibrium electron density (which will have a local net negative charge density) and will want to enter a region with lower than equilibrium electron density (which will have a local net positive charge density). This type of restoring force allows for the propagation of waves in the system, which is associated with bulk plasmons.

The word "plasmons" will be used frequently throughout this work. In reality, the waves described above could be more accurately identified as plasma waves. Plasmons would then be the quantization of these waves (which is a term from quantum mechanics meaning the particle aspect, rather than the wave aspect, of these oscillations). In other words, the relationship between plasmons and plasma waves is more-or-less analogous to the relationship between photons and electromagnetic waves. That being said, throughout this text (and throughout the literature as a whole) the word plasmon is used to describe these waves, without any explicit interest in the quantum mechanics of the electron sea. A discussion of this can be found in some textbooks dealing with many-body theory [2], but is not of great importance to the understanding of most current applications.

For our purposes, surface plasmons (SP) are of more interest than bulk plasmons. A surface plasmon is a wave confined to the metallic interface, propagating in two dimensions. The surface plasmon resonant frequency is lower than the bulk plasmon frequency (by a factor of $\sqrt{2}$ for an infinite planar interface) due to the reduced dimensionality.

1.2 THE DIELECTRIC CONSTANT OF METALS

The role of the dielectric function is to model how light interacts with matter. This is a critically important problem. Optics would be very boring if it was only concerned with the propagation of light in a vacuum. A more formal phrasing for this problem would be, "how does an electromagnetic field oscillating at a particular frequency (monochromatic light) interact with a large collection of polarizable particles?" If the optical properties of the system are linear the answer to this question is applicable to much more than monochromatic light; the principles of Fourier analysis hold that any general wave packet of light can be expressed as a weighted sum of monochromatic plane waves.

The approach taken by many optics or electromagnetism texts [1],[3] is to treat the atoms in the system as harmonic oscillators (springs). In other words, the electron is treated as a small mass attached to the relatively huge (and more-or-less fixed) nucleus. Even if the nucleus is not really fixed – say it is a nitrogen atom at room temperature, bouncing around erratically with moles of other nitrogen atoms – the oscillating electromagnetic field itself does little to perturb the overall atomic motion. The resulting relative dielectric function, taken from Yariv [1], is as follows.

$$\varepsilon(\omega) = 1 + \frac{Ne^2}{m\varepsilon_0(\omega_0^2 - \omega^2 - i\nu\omega)} \tag{1.1}$$

In this expression, N is the electron number density, e and m are the electronic charge and mass, ω is the frequency of the light (the driving frequency), and ω_0 and ν are the effective resonant frequency and damping of the electronic oscillation.

In the above model, the electron and nucleus are described as an oscillating dipole, tightly bound by a spring-like bond. But in a metal the valence electrons are delocalized and are not strongly bound to any particular nucleus. In the context of the spring model, this means that the restoring force, which is proportional to ω_0^2, is very small. We also note that N can be grouped with all the physical constants (m, e, ε_0) to produce a term with the units of frequency.

$$\frac{Ne^2}{m\varepsilon_0} \equiv \omega_P^2 \tag{1.2}$$

This expression is the plasma frequency mentioned previously. For any electron gas, of course the charge and mass will be constant; hence, the density of conduction electrons is the primary factor which determines the plasmon frequency. The metallic dielectric function – which is the square of the index of refraction - then reduces to the following.

$$\varepsilon(\omega) = n^2(\omega) = 1 - \frac{\omega_P^2}{\omega(\omega + i\nu)} \tag{1.3}$$

At frequencies larger than the plasma frequency (shorter wavelength), the real part of the dielectric function tends to be positive. At frequencies smaller than the plasma frequency (longer wavelength), the real part of the dielectric function tends to be negative. A negative permittivity corresponds to an imaginary index of refraction. An imaginary index of refraction signifies decay in the direction of propagation, meaning the fields will decay evanescently into the medium. Plasmonic phenomena are most significant in metals which exhibit a strongly negative-real dielectric constant at optical frequencies. Silver and gold are especially of interest in plasmonics applications (plots of silver and gold dielectric functions are shown in Figure 1). In fact, the properties which make them valuable in terms of plasmon support are not entirely de-coupled

from the properties which make them valuable for ornamental purposes. A strongly negative dielectric constant results in high-reflectivity, which produces the luster characteristic of precious metals.

Metals can support surface plasmon waves due to the possibility of negative permittivity. This will be shown quantitatively in Chapter 2 and 3 in the analysis of a metallic-dielectric interface (a topic which is already well established, but important as a point of comparison for more complicated systems).

1.3 TRANSMISSION THROUGH SUB-WAVELENGTH APERTURES

One of the great surprises of nano-optics is that sub-wavelength slits in metals exhibit unexpectedly large transmission relative to their size. More generally, a slit or hole in a metallic film has an effective area which can be larger than the actual area – this ratio can be one measure of the collection or funneling efficiency of the aperture. The problem of diffraction from sub-wavelength holes in a thin, perfectly conducting metal was studied in a landmark paper by Hans Bethe, written in 1944 [4]. The conclusion of his analysis was that the power transmitted by such an aperture should decrease rapidly and monotonically as the ratio of hole size to wavelength increases.

For example, Bethe calculates the diffraction cross section for a plane wave (with a k-vector, k) incident onto a 2D square array of holes of radius a. Shown below is the result from his work [4].

$$A_{CS}^{(TM)} = \frac{64}{27\pi} k^4 a^6 \cos^2 \varphi \tag{1.4}$$

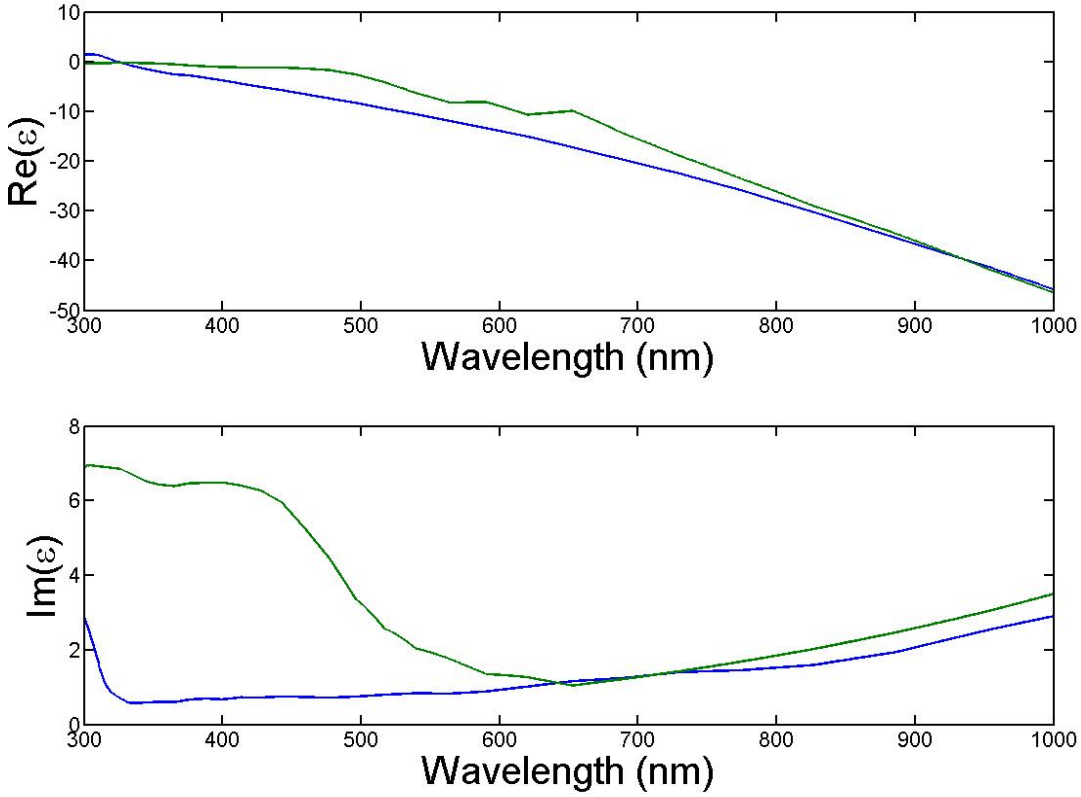

Figure 1: The dielectric function for silver (blue curve) and gold (green curve) over a range of wavelengths from near UV to near IR. The real parts of both are strongly negative at optical and IR wavelengths. Gold tends to have a much higher conductivity (proportional to Im(ε)) in the region from 300 to 600 nm and tends to be lossier in terms of plasmon support. Calculated using data from Johnson [5] and Palik [6].

$$A_{CS}^{(TM)} = \frac{64}{27\pi} k^4 a^6 \left[1 - \frac{3}{8} \sin^2 \varphi \right] \tag{1.5}$$

This result makes more sense if it is normalized by the actual area of the hole. In rough terms, this is analogous to comparing how much light falls on the hole area (A_H) to how much is actually transmitted.

$$\left(\frac{A_{CS}}{A_H} \right)_{TE} \propto k^4 a^4 \cos^2 \varphi \propto \left(\frac{a}{\lambda_0} \right)^4 \cos^2 \varphi \tag{1.6}$$

$$\left(\frac{A_{CS}}{A_H} \right)_{TM} \propto \left(\frac{a}{\lambda_0} \right)^4 \left[1 - \frac{3}{8} \sin^2 \varphi \right] \tag{1.7}$$

Here φ is the angle of incidence and TE is defined as the polarization state where the electric field is perpendicular to the plane of incidence (defined by the normal vector of the metal plane and the k-vector). The behavior of the two polarizations are similar in this case, due to the symmetry of the circular hole structure analyzed by Bethe. For the slit structure studied in the bulk of this thesis, the two polarizations will have significantly different properties.

It is not difficult to see that Bethe's model predicts that transmission should drop as wavelength increases for a given hole area. Experimental results on arrays of sub-wavelength holes in silver show much different transmission properties, along with observation of higher than unity normalized transmission efficiency [7]. In other words, if you picture light falling uniformly across a film covered with holes, intuition suggests that only the light falling on the holes can potentially be transmitted. Higher than unity transmission efficiency implies that light falling outside of the hole regions is being funneled into the holes. The fact that high

transmission tends to be observed in periodic structures is very suggestive, because periodicity is one of the components needed for light to be successfully coupled into surface plasmons.

There are three primary assumptions in his model. First, the wavelength is assumed to be much greater than the size of the hole. Second, the metal film is assumed to be very thin. It seems very unlikely that a thicker film would result in better transmission, so this is not a major issue. Finally, the metal is assumed to be a perfect conductor. This seems to suggest that the non-zero penetration depth of a real metal, along with the potential for negative permittivity may have something to do with the "extraordinary" transmission phenomenon. While the focus of this work will be on the role of plasmonic modes in sub-wavelength apertures, it should be pointed out that there is some contention regarding the effect of plasmon support on extraordinary transmission (specifically, it is claimed at extraordinary transmission can occur in perfectly conducting slits, which should not support plasmons) [8],[9]. In Section 2.4, one possible mechanism for enhanced transmission in metallic slits will be presented and backed up using finite-difference simulations. For a more in-depth literature review of light propagation in sub-wavelength slits, consult Genet and Ebbesen [10].

1.4 AN OVERVIEW OF APPLICATIONS

Many of the applications associated with surface plasmon phenomena utilize the strongly localized electromagnetic fields of SP waves. The energy associated with a SP propagating along a metal surface is mostly confined within a region of 100 nm or so perpendicular to the surface. The benefit of confinement sometimes comes from increased local energy density and sometimes purely from the reduced length scale. Many applications in traditional optics are restricted by the

diffraction limit of light, whereby light cannot be used to resolve dimensions less than about one wavelength. Plasmons offer a nice loophole, because the diffraction limit is significantly less strict in the near-field (on the order of a few wavelengths away from the structure). Of course, the capacity to design and perform experiments on the nanoscale is one of the pre-requisites for exploiting near-field phenomena. This is something which has only been feasible for the past decade or so.

Some interesting phenomena in surface physics are intrinsically related to the properties of surface plasmons. It have been widely established that the energy levels of a system can be shifted significantly by the close proximity of a metallic surface. One interesting application of this is that a fluorescent particle can be significantly enhanced or quenched (at very close distances, called Förster quenching) by strongly coupling it with a metallic interface [11],[12].

It will be shown analytically in Chapter 2 that the index of refraction observed by light propagating in a very narrow metallic slit can be relatively large. The index of refraction is roughly a measure of how strongly light can be bent and shaped. It has been proposed by Z. Sun et al. that an array of slits with varying thickness can exhibit some of the same properties as a lens, but on the nanoscale [13]. This device operates on essentially the same principles as a classical glass lens – the thickness of the lens is controlled to introduce a phase shift as a function of radius; the end result of adding up the phase-shifted waves is a sharply focused beam. Subsequent work by H. Shi et al. has illustrated that a similar effect can be achieved by varying slit width as a function of position [14]. The effective index of the slits is a strong function of slit width; if two slits are placed in a film of the same thickness, light leaving the smaller slit will exhibit a greater phase shift than the larger slit. The transmission properties of a grating can be a complex function of slit width and slit spacing, so design of such a system may be difficult.

Another application that will be discussed in Chapter 4 is the utilization of metallic layers in a distributed Bragg reflector (DBR) – DBRs are periodically alternating stacks of two or more different materials with different optical properties. The periodicity provides strong reflectivity in certain wavelength regions and strong transmission in other regions. The index of refraction in metals tends to be very large in magnitude; this is valuable because strong index contrast between layers is an important factor in the design of a DBR. While the index of refraction tends to be strongly imaginary - which should result in extremely high loss - the periodicity of this system produces some interesting and surprising results.

One of the most currently significant applications of SP waves is chemical sensing. The localized nature of surface plasmons makes them extremely sensitive to any modification of the surface upon which they are propagating. As a result, adding a thin layer of substance with an index of refraction slightly different than the background can substantially change the transmission properties of plasmonic systems. In practice, this is achieved by depositing a layer of a molecule (call it "A") which binds with the molecule you want to detect (call it "B"). That way, you can determine the base transmission of the device with molecule A deposited and the modified spectrum when A is saturated with B. If a device with A is then exposed to some environment which may or may not contain B, the change in the transmission spectrum reveals whether or not any B was caught by the surface. Many different systems have been devised to use surface plasmons for chemical sensing, with varying applicability and sensitivity: corrugated films, prisms, gratings, and so on. Our primary interest is in the 1D grating; a thorough overview of other methods and of plasmonic chemical sensing in general was written by Homola et al [15].

2.0 ELECTROMAGNETIC MODES IN A METALLIC SLIT

2.1 PROBLEM FORMULATION

This section will present the physical background information necessary to calculate the eigenmodes for a metallic slit waveguide, a dielectric slab separating two semi-infinite metallic regions. While this structure seems very simple, the results can be used to approximate more complicated systems. Moreover, the problem contains a number of physically interesting subtleties which will reveal much about light propagation in sub-wavelength slits. The geometry of interest is shown in Figure 2.

Like any problem in electromagnetics, Maxwell's equations will be the starting point. The time-frequency domain version will be utilized to remove any explicit time-dependence. The medium of interest is assumed to contain no bulk free charge.

$$\nabla \times \vec{H} = \vec{J} - i\omega\vec{D} \tag{2.1}$$

$$\nabla \times \vec{E} = i\omega\vec{B} \tag{2.2}$$

$$\nabla \cdot \vec{D} = 0 \tag{2.3}$$

$$\nabla \cdot \vec{B} = 0 \tag{2.4}$$

Each region under consideration is assumed to be linear and isotropic. The constituitive relations reduce the number of independent fields from four to two.

$$\vec{D} = \varepsilon(x)\varepsilon_0 \vec{E} \tag{2.5}$$

$$\vec{B} = \mu_0 \vec{H} \tag{2.6}$$

The dielectric constant is a function of x, and is constant in the y and z directions. If the y-direction is selected as the direction of propagation, all the fields will be constant with respect to z. Assuming that the magnetic field is parallel to the z-direction (referred to as the TM polarization), Maxwell's equations can be reduced to a scalar wave equation with two boundary conditions at each interface.

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \varepsilon\mu\left(\frac{\omega}{c}\right)^2\right] H_z(x,y) = 0 \tag{2.7}$$

$$H_z(x^+_{Interface}, y) = H_z(x^-_{Interface}, y) \tag{2.8}$$

$$\frac{1}{\varepsilon(x^+_{Interface})}\frac{\partial H_z}{\partial x}(x^+_{Interface}, y) = \frac{1}{\varepsilon(x^-_{Interface})}\frac{\partial H_z}{\partial x}(x^-_{Interface}, y) \tag{2.9}$$

The expression in 2.8 requires the continuity of the component of the magnetic field tangential to the interface (H_z) and 2.9 requires the continuity of the component of the electric field tangential to the interface (E_y). The solutions will also assumed to be bounded, requiring that all fields go to zero far away from the slit region. Due to the large imaginary part of the index of refraction, light dies off exponentially in a metal above the plasmon frequency; so boundedness is a safe assumption.

$$\lim_{x \to \pm\infty} H_z(x,y) = 0 \tag{2.10}$$

$$\lim_{x \to \pm\infty} \frac{\partial H_z(x,y)}{\partial x} = 0 \tag{2.11}$$

The solutions to the wave equation will be harmonic; as a result, any valid solution can be built up as a sum of complex exponentials.

$$H_z(x,y) = \sum_m H_m \exp\left(i\vec{k}_m \cdot \vec{r}\right) = \sum_m H_m \exp\left(i\beta^{(m)} y\right)\exp\left(ik_x^{(m)} x\right) \tag{2.12}$$

The component of the wave vector in the y-direction is called β and (for now) the component in the x-direction will be designated k_x. Of course, all solutions must satisfy 2.7.

$$\left[-k_x^2 - \beta^2 + \varepsilon\mu\left(\frac{\omega}{c}\right)^2\right] H_z(x,y) = 0 \tag{2.13}$$

In some regions, the fields could potentially be evanescent in the x-direction. In other words, k_x could be imaginary, changing 2.13 slightly.

$$\left[\gamma^2 - \beta^2 + \varepsilon\mu\left(\frac{\omega}{c}\right)^2\right] H_z(x,y) = 0 \tag{2.14}$$

Here γ is simply defined as ik_x. There is one final relationship that will be useful in finding explicit expressions for the fields. It can be easily seen that the geometry is symmetric with respect to the center of the slit, therefore the fields must have similar symmetry.

$$H_z(0,y) = \pm H_z(a,y) \tag{2.15}$$

All of the physics of the problem has been summarized in a few mathematical statements. These will be the building blocks for the overall expressions for the fields. From here, constructing the fields is a more or less straightforward process. In the next section, the bound state solutions for the fields will be presented without explicit derivation, so that more interesting aspects of the system can be emphasized.

2.2 GENERAL EIGENVALUE PROBLEM

In the analysis of surface plasmon waves, there are a few basic features which can be guessed from physical intuition alone. Assuming that the system supports surface waves, the electromagnetic energy associated with that wave is expected to be bound to the surface – that is, an exponential drop off in the fields in the direction normal to the surface is expected. It should be pointed out that there is no loss of generality in this assumption. If no surface waves can be supported, the analysis will correct our pre-conceptions.

The simplest possible structure which can support surface plasmon waves is a single metallic interface, where a semi-infinite metal region meets a semi-infinite dielectric region. A SP propagating along such an interface will travel at a speed slightly less than the speed of light. The electric and magnetic fields associated with a single interface are derived by Raether [16]. The decay constants on each side of the interface and the propagation constant along the surface are as follows.

$$\beta = k_0 \sqrt{\frac{\varepsilon_M \varepsilon_D}{\varepsilon_M + \varepsilon_D}} \tag{2.16}$$

$$\gamma_D = \sqrt{\beta^2 - \varepsilon_D k_0^2} \tag{2.17}$$

$$\gamma_M = \sqrt{\beta^2 - \varepsilon_M k_0^2} \tag{2.18}$$

The relative dielectric constants of the metal and dielectric sides are denoted by ε_M and ε_D, and k_0 is the free-space wave vector ($k_0 = 2\pi/\lambda_0 = \omega/c$). The inverse of γ_D (or γ_M) gives the 1/e decay length for the fields on the dielectric (or metal) side. For typical systems of interest, $\varepsilon_M < 0$ and $|\varepsilon_D| < |\varepsilon_M|$. From the expressions above, this also implies that $\gamma_D < \gamma_M$. This indicates greater penetration on the dielectric side.

The single isolated slit structure (SIS) is realized by bringing two metallic interfaces close together; this results in coupling between the waves at the two interfaces. A complementary structure would be a two-dimensional metal slab of finite thickness [17]. The fields associated with each interface can interfere with each other symmetrically (constructively) or anti-symmetrically (destructively).

By bringing two interfaces close together, the fields no longer have infinite distance to decay into on the dielectric side. Equation 2.17 suggests that by decreasing the spatial extent on the dielectric side (by forcing γ_D to increase), β should increase. In other words, by confining the fields on one side the propagation constant increases. Physically this means that the effective index of refraction that the mode feels is increased.

It is well-established that the TE field does not propagate in this sort of structure, so we will analyze only the TM case. It will be shown in the next chapter that surface plasmon support for the TM mode requires a sign change in the dielectric constant across an interface; likewise, a TE surface wave *could* be supported if there was a sign change in the magnetic permeability across an interface. While potentially interesting, magnetic materials are not our primary interest.

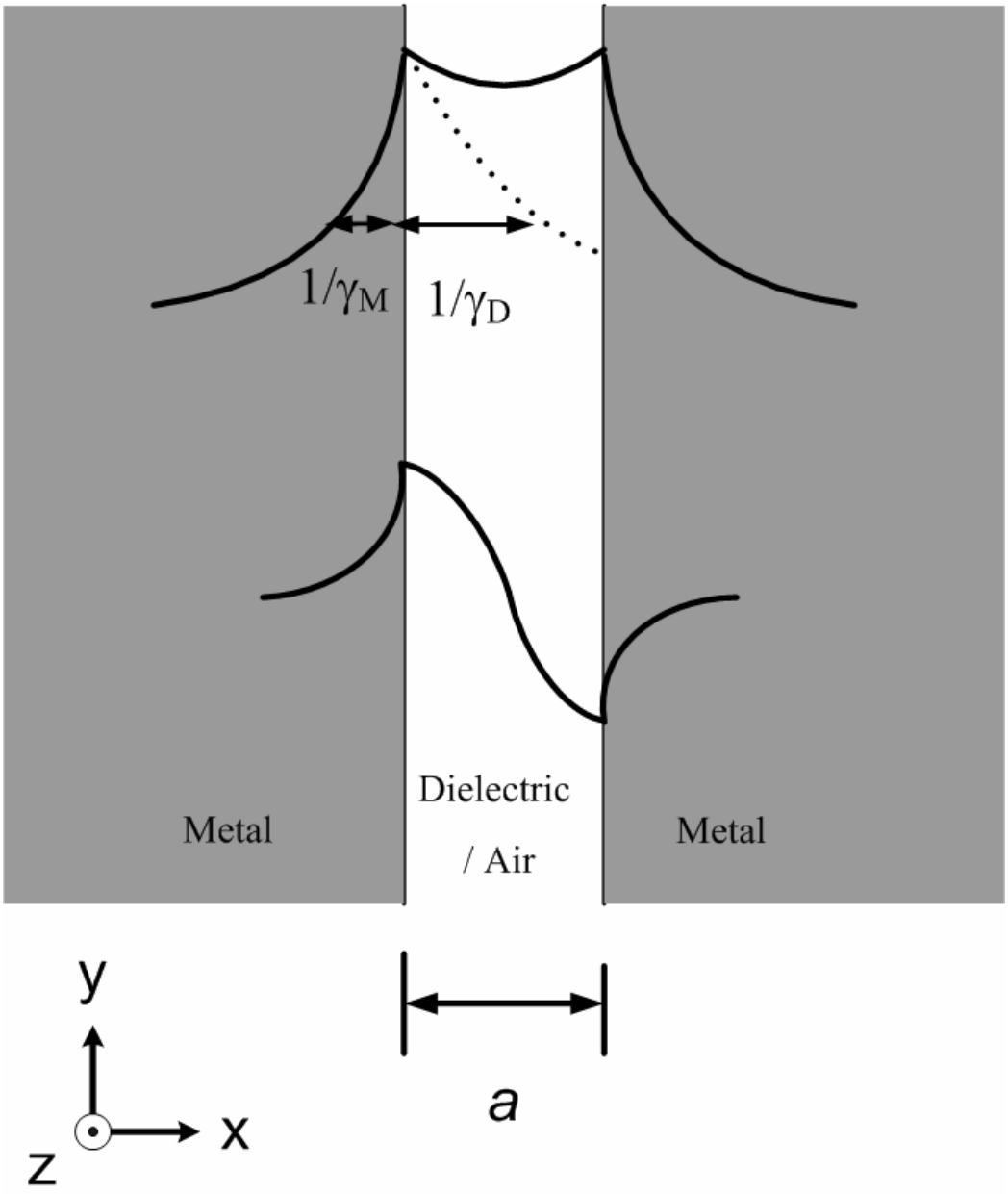

Figure 2: A rough sketch of the magnetic field (H_z) for the '+' solution to 2.19 – 2.27 (deemed the symmetric surface plasmon – SSP – mode, referring to the symmetry of the magnetic field) and the first order wave-guide like mode. The fields are primarily confined to the dielectric region (designated with ε_D). Note that a real value of γ_D is assumed. Propagation occurs in the y-direction.

After applying Maxwell's equations and the relevant boundary conditions (continuity of the tangential electric and magnetic fields), the results are as follows (in SI units).

$$H_z^{(1)}(x,y) = H_0(1 \pm e^{-\gamma_D a})e^{\gamma_M x}e^{i\beta y}, \text{ x} < 0 \tag{2.19}$$

$$H_z^{(2)}(x,y) = H_0(e^{-\gamma_D x} \pm e^{-\gamma_D (a-x)})e^{i\beta y}, \text{ 0} < \text{x} < \text{a} \tag{2.20}$$

$$H_z^{(3)}(x,y) = H_0(e^{-\gamma_D a} \pm 1)e^{-\gamma_M (x-a)}e^{i\beta y}, \text{ x} > \text{a} \tag{2.21}$$

$$E_y^{(1)}(x,y) = \frac{\eta_0}{i\varepsilon_M}\left(\frac{\gamma_M}{k_0}\right)H_0(1 \pm e^{-\gamma_D a})e^{\gamma_M x}e^{i\beta y}, \text{ x} < 0 \tag{2.22}$$

$$E_y^{(2)}(x,y) = -\frac{\eta_0}{i\varepsilon_D}\left(\frac{\gamma_D}{k_0}\right)H_0(e^{-\gamma_D x} \mp e^{-\gamma_D (a-x)})e^{i\beta y}, \text{ 0} < \text{x} < \text{a} \tag{2.23}$$

$$E_y^{(3)}(x,y) = -\frac{\eta_0}{i\varepsilon_M}\left(\frac{\gamma_M}{k_0}\right)H_0(e^{-\gamma_D a} \pm 1)e^{-\gamma_M (x-a)}e^{i\beta y}, \text{ x} > \text{a} \tag{2.24}$$

$$E_x^{(1)}(x,y) = -\frac{\eta_0}{\varepsilon_M}\left(\frac{\beta}{k_0}\right)H_0(1 \pm e^{-\gamma_D a})e^{\gamma_M x}e^{i\beta y}, \text{ x} < 0 \tag{2.25}$$

$$E_x^{(2)}(x,y) = -\frac{\eta_0}{\varepsilon_D}\left(\frac{\beta}{k_0}\right)H_0(e^{-\gamma_D x} \pm e^{-\gamma_D (a-x)})e^{i\beta y}, \text{ 0} < \text{x} < \text{a} \tag{2.26}$$

$$E_x^{(3)}(x,y) = -\frac{\eta_0}{\varepsilon_M}\left(\frac{\beta}{k_0}\right)H_0(e^{-\gamma_D a} \pm 1)e^{-\gamma_M (x-a)}e^{i\beta y}, \text{ x} > \text{a} \tag{2.27}$$

In this case, the expressions for γ_D and γ_M in 2.17 and 2.18 are still valid (they are direct results of the wave equation), but the propagation constant changes. The eigenvalue equation for the propagation constant follows directly from the boundary conditions.

$$\frac{\gamma_D}{\varepsilon_D}\tanh^{\pm 1}\left(\frac{\gamma_D a}{2}\right) = -\frac{\gamma_M}{\varepsilon_M} \tag{2.28}$$

In order to determine the propagation constant, this expression must be solved numerically in terms of β (which appears implicitly through γ_D and γ_M). One straightforward way of solving 2.28, is to define the following function.

$$F_{\pm}(\beta) = \left| \frac{\gamma_D(\beta)}{\varepsilon_D} \tanh^{\pm 1}\left(\frac{\gamma_D(\beta) a}{2} \right) + \frac{\gamma_M(\beta)}{\varepsilon_M} \right| \geq 0 \tag{2.29}$$

The eigenvalue equation is obviously satisfied when $F_{\pm}$ is equal to zero – its minimum possible value. The solution (or solutions) for β can then be found by numerically locating local minima of $F_{\pm}$. In general, this will be a two dimensional minimization problem, the two "independent" variables being the real and imaginary parts of β.

At the beginning of the derivation, the goal was explicitly to find plasmon-like modes, where the fields decay evanescently away from the interface. In order for this to be true, the expressions in 2.17 and 2.18 have to result in real numbers. There is of course no physical reason why the fields cannot be oscillatory in the dielectric region. In fact, the usual standing wave solutions in a perfect electric conductor (PEC) waveguide or the approximate standing wave solutions in a dielectric slab waveguide are perfect examples of this type of solution. So, let us see what happens if γ_D is allowed to be imaginary.

$$\gamma_D \equiv i k_D \tag{2.30}$$

$$i \frac{k_D}{\varepsilon_D} \tanh^{\pm 1}\left(i \frac{k_D a}{2} \right) = -\frac{\gamma_M}{\varepsilon_M} \tag{2.31}$$

To simplify this, the following relationship is utilized (which can be easily derived by writing the tangent and hyperbolic tangent as complex exponentials).

$$\tanh^{\pm 1}(ix) = \pm i \tan^{\pm 1}(x) \tag{2.32}$$

$$\mp \frac{k_D}{\varepsilon_D} \tan^{\pm 1}\left(\frac{k_D a}{2}\right) = -\frac{\gamma_M}{\varepsilon_M} \tag{2.33}$$

This version of the eigenvalue equation reveals that the imaginary solutions to 2.28 are the same as the real solutions to the traditional waveguide eigenmode equation [18]. To further illustrate this point, consider the solutions for a perfect electric conductor (PEC). The electric field must be zero everywhere in a PEC in steady state, so the penetration into the metal goes to zero. This can be achieved by letting the magnitude of the dielectric constant go to infinity on the right-hand side of 2.28.

$$\lim_{|\varepsilon_M| \to \infty} \frac{\gamma_M}{\varepsilon_M} = \lim_{|\varepsilon_M| \to \infty} \frac{\sqrt{k_{sp}^2 - |\varepsilon_M| e^{i\varphi_M} k_0^2}}{|\varepsilon_M| e^{i\varphi_M}} = \lim_{|\varepsilon_M| \to \infty} \frac{ik_0}{\sqrt{|\varepsilon_M|}} e^{-i\frac{\varphi_M}{2}} = 0 \tag{2.34}$$

Note that this limit is independent of the phase of the metallic dielectric constant, φ_M. Physically speaking, a PEC is an idealized material with infinite conductivity; in other words, it is modeled with an infinite imaginary dielectric constant ($\varphi_M = \pi/2$). The result in 2.34 illustrates that the modal eigenvalues will be the same regardless of the phase of the dielectric constant. This is interesting because at wavelengths significantly longer than the plasmon resonance wavelength, the dielectric constant becomes a very large, but mostly real, negative number. As long as the magnitude of the dielectric constant is large, the results should be approximately equal to the results for the PEC waveguide. This allows 2.33 to be simplified for the PEC limit.

$$\mp \frac{k_D}{\varepsilon_D} \tan^{\pm 1}\left(\frac{k_D a}{2}\right) = 0 \tag{2.35}$$

$$k_D^+ \sin\left(\frac{k_D^+ a}{2}\right) = 0 \tag{2.36}$$

$$\cos\left(\frac{k_D^- a}{2}\right) = 0 \tag{2.37}$$

These will be standing wave solutions. In other words, the solutions will satisfy the requirement that the round-trip phase shift will be equal to multiples of π radians.

$$k_D a = m\pi\ , \qquad m = 0,1,2... \tag{2.38}$$

$$N_D = \frac{m}{2}\left(\frac{\lambda_0}{a}\right), \qquad m = 0,1,2... \tag{2.39}$$

This is a relatively simple and well established result from classical waveguide theory. It is presented here as a point of comparison. Theoretically, by simply mathematically 'turning up' the dielectric constant, the result for a realistic metal can be smoothly transformed into the PEC result. Intuitively, there should be a one-to-one relationship between the PEC modes and the sub-wavelength metallic waveguide modes. This idea will be discussed at length in the next section.

2.3 PLASMONIC AND WAVEGUIDE-LIKE MODES

Considering a particular metallic slit system, there are essentially two physical parameters which determine the propagation constant: the frequency of the light (directly related to the free-space wavelength) and the slit width. It can be seen from the PEC waveguide result in 2.39 that the effective index scales perfectly with λ_0/a. The general result is not nearly as simple, due to the

introduction of another important length scale in the system – the resonant plasma wavelength ($\lambda_P = c/2\pi\omega_P$) of the dielectric constant. Using the eigenvalue equation in 2.28 for a silver slit filled with air ($\varepsilon_D = 1$) and realistic values for the complex dielectric function of silver, the real and imaginary parts of the propagation constant for the first few modes at 650 nm ($\varepsilon_M = -17.0 + i1.15$, from Palik [6]) were calculated over a range of slit widths and are shown in Figure 3.

At long wavelengths, it can be seen from Figure 1 that the magnitude of the silver dielectric constant is very large. For this reason, it is reasonable to expect that the modes will be very similar to the PEC solutions. In order to correctly match these to the zeroes of 2.29, the symmetry of the modes must be correctly identified.

m: odd $\rightarrow$ E_y is symmetric $\rightarrow$ H_z is anti-symmetric
m: even $\rightarrow$ E_y is anti-symmetric $\rightarrow$ H_z is symmetric

If the desired mode is symmetric, the positive function given in 2.29 should be minimized. Because the result for k_D (γ_D) is simple and known in the long-wavelength limit (k_D is approximately equal to $m\pi/a$ for the $m > 0$ modes) it is convenient to pose the minimization problem in terms of this variable.

$$\left| \frac{\gamma_D}{\varepsilon_D} \tanh\left(\gamma_D \frac{a}{2} \right) + \frac{\gamma_M(\gamma_D)}{\varepsilon_M} \right| = F^{(+)}(\gamma_D) \geq 0 \tag{2.40}$$

$$\gamma_M(\gamma_D) = \sqrt{\gamma_D^2 + (\varepsilon_D - \varepsilon_M)k_0^2} \tag{2.41}$$

Anti-symmetric modes can be solved similarly.

$$\left| \frac{\gamma_D}{\varepsilon_D} \coth\left(\gamma_D \frac{a}{2} \right) + \frac{\gamma_M(\gamma_D)}{\varepsilon_M} \right| = F^{(-)}(\gamma_D) \geq 0 \tag{2.42}$$

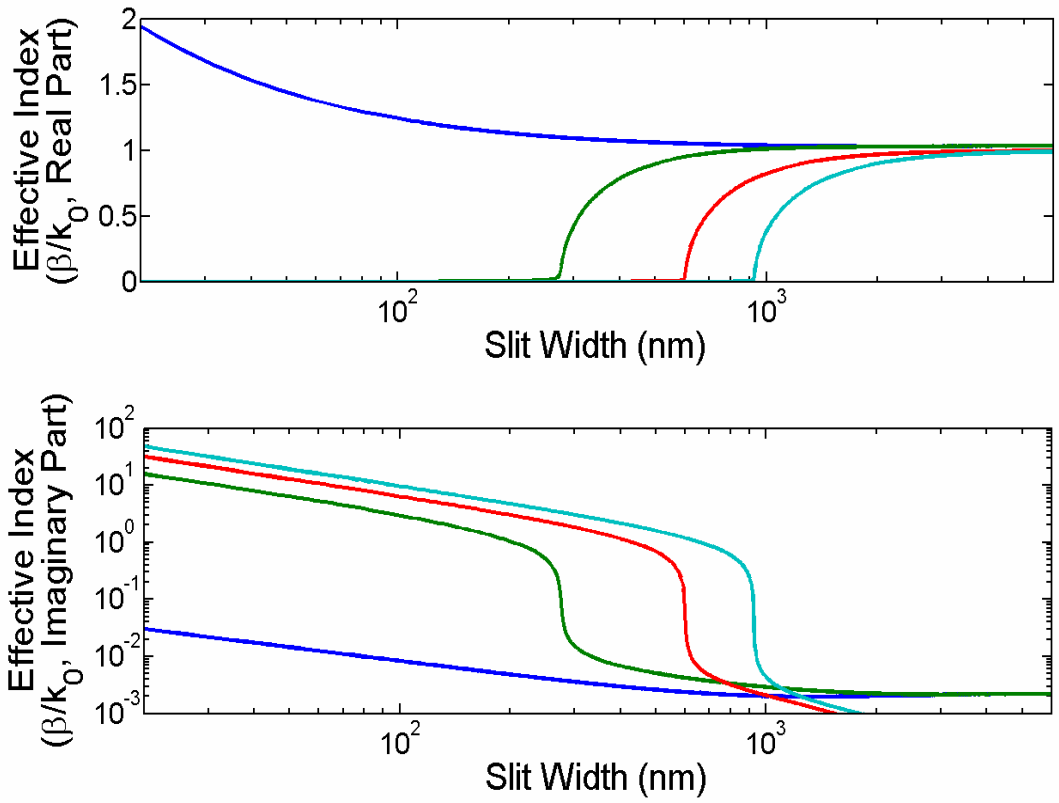

Figure 3: The effective index as a function of slit width (ranging from 6000 nm down to 20 nm) at 650 nm free-space wavelength. The symmetric SP mode (blue), the anti-symmetric SP mode / $m = 1$ WG mode (green), the $m = 2$ WG mode (red), and the $m = 3$ WG mode (teal) are shown.

A convenient initial guess for the solution of the symmetric and anti-symmetric plasmon modes at large slit widths is the completely decoupled result for a single metallic interface.

$$\gamma_{D,isolated} = k_0 \frac{\varepsilon_D}{\sqrt{-\varepsilon_M - \varepsilon_D}} \tag{2.43}$$

For the WG modes, the perfect conductor result is used as the initial guess.

$$\gamma_D = im\frac{\pi}{a} \tag{2.44}$$

The value of γ_D can then be solved as a two dimensional minimization problem (in the complex plane) over a range of slit widths. For most of the values the initial guess at a slit width "a" is the calculated solution at the previous step "$a + \Delta a$". Because effective index is a continuous function, the quality of this guess is very good at small values of Δa. This method is a little trouble-some in regions where two modes have close to the same eigenvalues; if the solver accidentally jumps to the wrong mode, this incorrect result will seed the initial guess for the new result, and it will stay on that mode. Figure 3 seems to show that all the modes are very close at long slit widths, so the physically based guesses described in 2.43 and 2.44 are mostly used in this region.

The results for the m > 1 WG modes are as expected. Due to the extremely high dielectric constant of the metal, larger field penetration on the dielectric side of the interface is energetically preferable. When the slit width is large, there is no confinement effect on the dielectric side. For the waveguide-like modes, this means they can spread out all the way to the other side of the slit –

however far that might be – and most of the power will be located in the dielectric region far away from the interface. As a result, the effective index is approximately equal to $\varepsilon_D^{1/2}$ =1. The plasmonic modes are different; the wave is naturally bound to the surface of the metal-dielectric interface. So, at even very large slit widths, the penetration on the air side is limited. For frequencies of interest, ε_M is negative and $|\varepsilon_M| > \varepsilon_D$. Judging from the result for an isolated interface in 2.16, the effective index is expected to be slightly greater than one in the limit as slit width goes to infinity (for both the symmetric and anti-symmetric modes).

Something interesting happens to the m = 1 mode at long slit widths. For a PEC waveguide, the m = 1 mode would – as discussed above – go to N = 1 in the long-slit width limit. But the numerical results show that above a certain slit width, the effective index goes greater than one and actually converges to the single-interface plasmonic solution. In the same way, the effective index solution initially assumed to be associated with the anti-symmetric plasmon mode eventually drops below unity and converges with the m = 1 waveguide like mode. The interpretation of this is that they are actually the same mode; the γ_D solutions for this mode are real above a certain slit width and imaginary below this slit width. This transition point can be estimated by finding where γ_D^2 crosses zero (goes from positive to negative). Using 2.17, 2.18 and the eigenvalue equation, the transition slit width can be determined from a few steps of algebra.

$$\beta = \sqrt{\varepsilon_D} k_0 \tag{2.45}$$

$$\gamma_M = \sqrt{\varepsilon_D - \varepsilon_M} k_0 \tag{2.46}$$

$$\lim_{\gamma_D \to 0} \frac{\gamma_D}{\varepsilon_D} \coth\left(\frac{\gamma_D a}{2}\right) = -\frac{\sqrt{\varepsilon_D - \varepsilon_M}}{\varepsilon_M} k_0 \tag{2.47}$$

$$\lim_{\gamma_D \to 0} \frac{\gamma_D}{\varepsilon_D} \frac{e^{\frac{\gamma_D a}{2}} + e^{-\frac{\gamma_D a}{2}}}{e^{\frac{\gamma_D a}{2}} - e^{-\frac{\gamma_D a}{2}}} = -\frac{\sqrt{\varepsilon_D - \varepsilon_M}}{\varepsilon_M} k_0 \tag{2.48}$$

$$\frac{\gamma_D}{\varepsilon_D} \frac{2}{\gamma_D a} = -\frac{\sqrt{\varepsilon_D - \varepsilon_M}}{\varepsilon_M} k_0 \tag{2.49}$$

$$a_{Trans} = \left(-\frac{\varepsilon_M}{\varepsilon_D} \right) \frac{1}{\sqrt{\varepsilon_D - \varepsilon_M}} \frac{\lambda_0}{\pi} \tag{2.50}$$

Physically speaking, the m = 1 mode is sinusoidal in the slit region below this slit width and is evanescently surface-bound above this slit width. In the PEC limit discussed above – where $|\varepsilon_M|$ goes to infinity – it can be easily shown that this transition slit width goes to infinity as $|\varepsilon_M|^{1/2}$. The PEC solutions should be sinusoidal standing waves in the slit region for *all* slit widths, so this limit makes sense.

All four modes appear to grow without bound at small slit widths. The small slit width effective index can be approximated by letting β dominate in 2.17 and 2.18 ($|N| >> |\varepsilon_M|^{1/2}$).

$$\frac{N}{\varepsilon_D} k_0 \tanh^{\pm 1}\left(\frac{N k_0 a}{2} \right) = -\frac{N}{\varepsilon_M} k_0 \tag{2.51}$$

$$\tanh^{\pm 1}\left(N\pi \frac{a}{\lambda_0} \right) = -\frac{\varepsilon_D}{\varepsilon_M} \equiv \alpha \tag{2.52}$$

The ratio of the dielectric constants is temporarily defined as α for notational simplicity. It is also convenient to temporarily define the argument of the hyperbolic tangent as x.

$$\tanh^{\pm 1}(x) = \left(\frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \right)^{\pm 1} = \alpha \tag{2.53}$$

After a few steps of algebra, an explicit expression for e^x can be derived.

$$e^{2x} = \frac{1+\alpha^{\pm 1}}{1-\alpha^{\pm 1}} = \pm\left(\frac{1+\alpha}{1-\alpha}\right) \tag{2.54}$$

$$2x = 2N\pi\frac{a}{\lambda_0} = \ln\left[\pm\left(\frac{1+\alpha}{1-\alpha}\right)\right] = \ln(\pm 1) + \ln\left[\left(\frac{1+\alpha}{1-\alpha}\right)\right] \tag{2.55}$$

For the waveguide modes at small slit widths, the effective index is a complex number with a large imaginary part. The natural log is not single-valued everywhere on the complex plane; this is a subtle point, but is very important in this context.

$$\ln(\pm 1) \rightarrow \ln(e^{i\pi m}) = i\pi m \tag{2.56}$$

Where the even values of m correspond to ln(+1) and the odd values correspond to ln(-1). Using this property of the natural log, substituting the expression for a back in, and performing a few steps of algebra, a complex expression for the effective index at small slit widths falls out.

$$N = i\frac{m}{2}\left(\frac{\lambda_0}{a}\right) + \frac{1}{2\pi}\left(\frac{\lambda_0}{a}\right)\ln\left(\frac{\varepsilon_M - \varepsilon_D}{\varepsilon_M + \varepsilon_D}\right),\ m = 0,1,2,3,... \tag{2.57}$$

Here $m = 0$ is the result for the symmetric surface plasmon mode. This result requires some qualification. First, it is only accurate for the SSP mode at very small slit widths, simply because

|N| grows slowly compared to the other modes in the limit as slit width goes to zero. Also, it does not predict the nearly zero real part of the effective index for the waveguide-like modes. The imaginary part is so large, however, that the relative error is not significant (both the magnitude and complex phase of the effective index will be fairly accurate). This arrives from the assumption that the magnitude of N is large – no attention was paid to the individual real and imaginary parts, so there is some sloppiness involved in calculating the real part of the result.

Arguably, the most important result from Figure 3 is that the symmetric SP mode (m = 0) – and only the symmetric SP mode – is not cut off at small slit widths. This result is supported by the expression in 2.57; the inherently imaginary part of the effective index is zero when m = 0. The loss imposed on the m = 0 mode is non-zero only when the dielectric constant of the metal is treated as complex. In other words, the m = 0 mode attenuates only due to the lossiness of the metal, whereas the loss of the metal is insignificant for the m > 0 modes compared to the attenuation caused by small slit width cutoff. Although the loss associated with the symmetric SP mode grows as the slit width decreases – a symptom of greater field penetration being pushed into the lossy metal region due to confinement on the air side – it is orders of magnitude less than the loss associated with the m > 0 modes, as illustrated by the imaginary component of the effective index. In fact, the imaginary component of the effective index is virtually neglible except at very small slit widths; from an engineering perspective, this regime is not particularly interesting, because it is currently difficult to reliably fabricate slit structures with slit width significantly less than 10 nm.

The symmetric SP mode also stands out because it is the only mode which remains surface-bound at small slit widths. This is ideal for chemical sensing applications for two reasons. First, most methods of plasmonic sensing utilize the strong-surface confinement of the

electromagnetic power. The SP mode would be especially sensitive to chemical surface modification, because it is localized in the region near the interface. Second, the effective index is significantly more sensitive in the small-slit regime simply due to the enhanced confinement of the modes.

To help illustrate the effect of the dielectric constant and how the solutions evolve from the PEC result to a solution for a more reasonable metal, the field profiles (specifically for the y-component of the Poynting vector; this is chosen because it best illustrates the power flow in the direction of propagation, to be discussed in full detail in the next chapter) for the m = 0, 1, 2 and 3 modes at 500 nm free-space wavelength and slightly under 600 nm slit width are plotted in Figure 4 as the metallic dielectric constant is relaxed from negative infinity (PEC equivalent) to $\varepsilon_M = -5$. The TM_1 mode is perhaps the most interesting, because the gradual transition from sinusoidal variation to exponential variation in the slit region can be observed as the magnitude of ε_M is decreased.

2.4 COUPLING AND TRANSMISSION

In the previous two chapters, heavy emphasis was placed on the analysis of modes inside a metallic slit. It is very important – but more difficult – to study the electrodynamics of light entering and leaving the slit. This will influence how light is transmitted and coupled into a metallic slit array.

In traditional waveguide analysis, coupled mode theory (CMT) can be used to predict how much power is coupled into a guide from a particular excitation source [1],[19]. CMT is based on

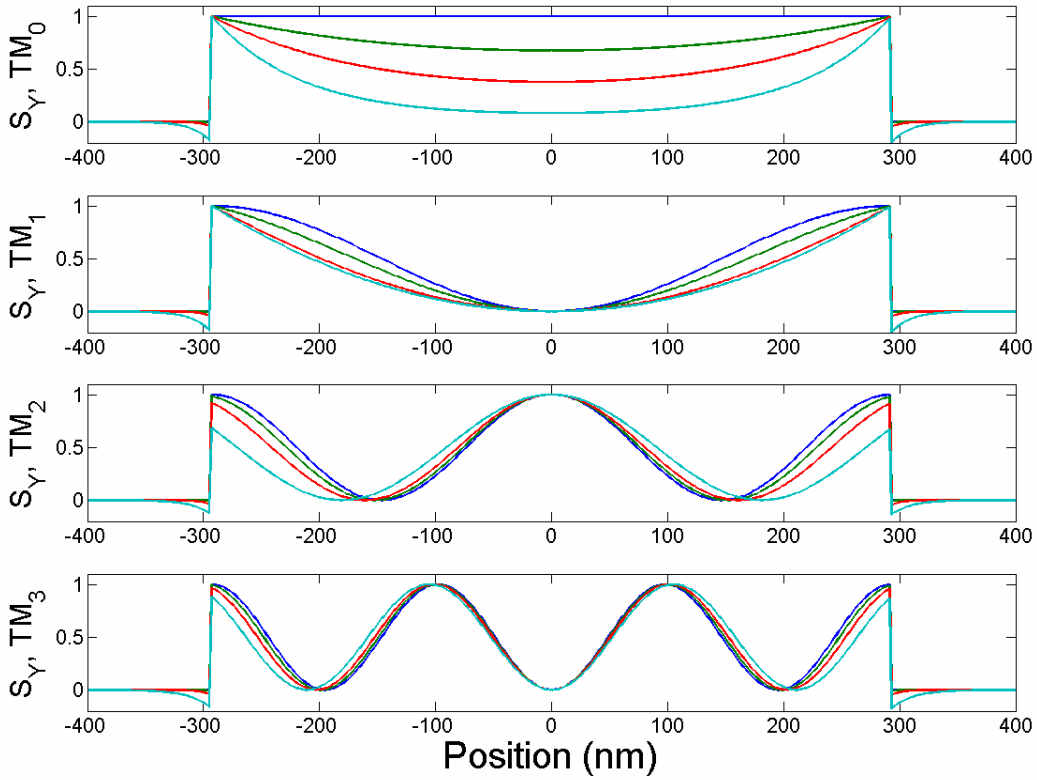

Figure 4: The field profiles at 500 nm free-space wavelength for the lowest four TM modes as the PEC condition is relaxed. The blue curve shows the PEC result, the green shows $\varepsilon_M = -100$, the red shows $\varepsilon_M = -20$ and the teal shows $\varepsilon_M = -5$

the premise that the waveguide modes form a complete set; in other words, any excitation source in the region of the guide entrance can be decomposed into a sum of the eigenmodes. Then the distribution of the power between the modes inside the guide will be the same as their relative contributions to the excitation source.

There are a number of subtle aspects to CMT that make it a little unwieldy for the problem at hand. Primary is the fact that a significant amount of power is coupled to SPs on the front of the metal film. This power flow is approximately normal to the incident wave and one would expect the overlap integral between them to be zero; however, it will be shown that the sharp spatial discontinuity associated with the slit edges allows for this seeming contradiction. In general, CMT does not allow for power transfer between modes if they are truly orthogonal. Also, due to loss, the Poynting vector associated with the SP modes are not entirely normal to the incident wave, so CMT could potentially be modified to account for these modes.

Numerical methods are useful in situations where analytic results are not readily available. One of the most widely used numerical methods for electromagnetic problems is the finite-difference time domain method (FDTD). In FDTD, the electric and magnetic fields are calculated over a discrete mesh and the differential versions of Maxwell's equations are approximated as a set of difference equations. There are many different approximations that can be used; the complexity of the algorithm needed varies greatly depending on the geometry and the system of interest. A brief discussion of the FDTD analysis of metallic systems is given in Appendix A.

By starting with a very simple system and slowly adding complexity, an intuitive picture of how light interacts with a metallic grating can be built. Consider first a single metal "corner" (i.e., a 90º wedge); in this simulation, the upper-left quadrant of the problem space is filled with

silver and the rest is filled with air. It is generally accepted that sharp corners play an important role in the generation of surface plasmons. There is a significant momentum mismatch between free light and SPs; a sharp spatial discontinuity provides the Fourier components necessary to convert free light to SPs and conserve momentum. A plane wave at 650 nm free-space wavelength is propagating in the +y-direction from the bottom of the problem space. After the system is allowed to reach steady-state (after the light has propagated all the way to the end of the problem space), the program begins calculating the time-averaged Poynting vector. The Poynting vector is averaged over approximately 1000 time steps, which is significantly longer than a single period (around 130 time steps). The x and y components of the time-averaged Poynting vector are shown in Figure 5. Red indicates positive values and blue indicates negative values; in other words, in red areas on the $<S_x>$ plot red indicate that power is flowing to the right (+x direction) and blue areas indicate that power is flowing to the left (-x direction).

In the plot of the y-component of the Poynting vector, strong forward (+y) power flow is seen near the right-hand metal-air interface. As expected, this seems to indicate that surface plasmons are generated at the corner and travel away from it. Further evidence is given by the large blue region on the front face near the corner in the $<S_x>$ plot. This represents strong power flow in the –x direction near the edge, suggesting that SPs on the front face are also propagating away from the corner. There is a rather complicated but weak circulation of power on the front face. This seems to be caused by interference between the many interacting waves in the region: the incident light, the light reflected from the front face, the light scattered from the corner, and and the fields associated with the SPs confined to that interface.

One of the most interesting aspects of this plot is the very small red region on the $<S_x>$ plot right next to the corner. As stated above, the larger blue region near the edge indicates

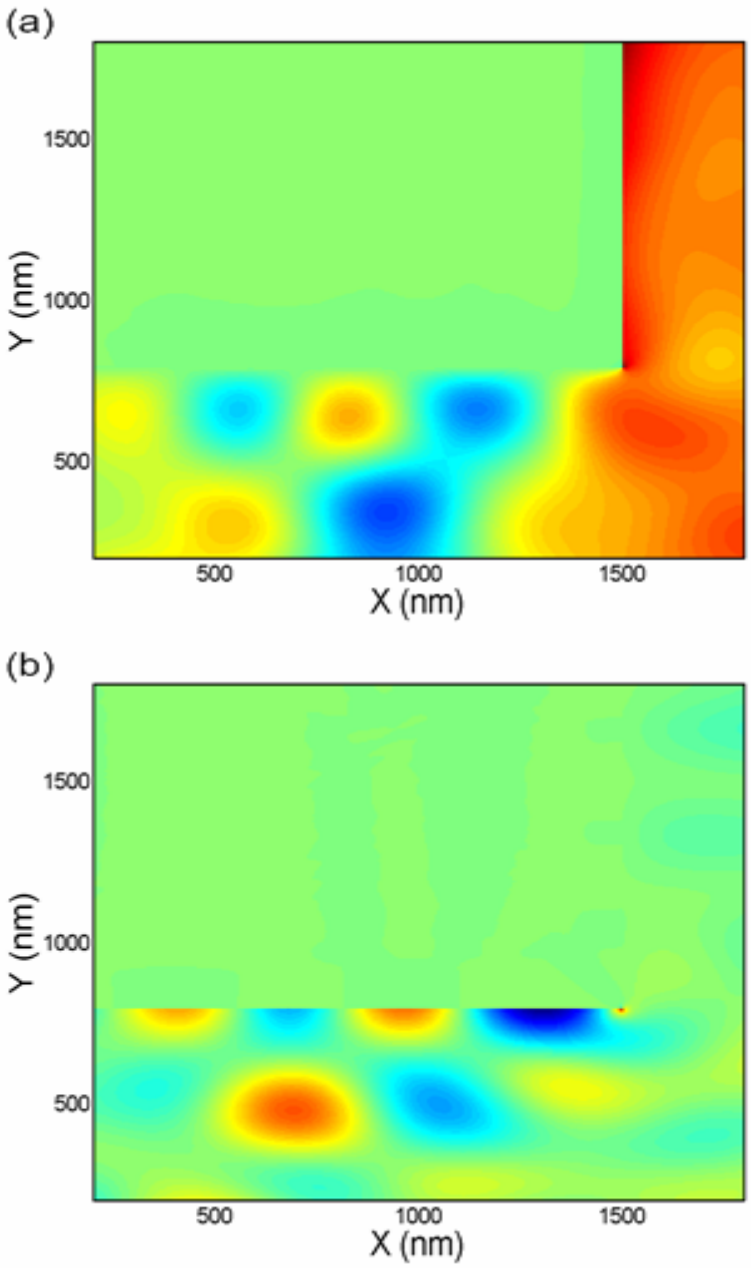

Figure 5: Time-averaged Poynting vector of plane wave (λ_0 = 650 nm) traveling in the +y-direction incident on a 90º silver corner. The (a) y-component and the (b) x-component of the Poynting vector are shown separately.

significant power flow away from the corner on the front face. But the red region reveals that very close to the edge (less than 100 nm), power is essentially flowing around the corner.

The near-field results for the FDTD analysis of a 200 nm thick silver film with a number of 100 nm wide slits (300 nm slit spacing) are shown in Figure 6. The system is excited with normally incident plane waves, at a free-space wavelength of λ_0 = 650 nm. The time-averaged Poynting vector is shown for systems with one, two and three slits.

There a number of interesting conclusions that can be drawn from the time-averaged Poynting vector plots in Figure 6. Similar to the simulation of the isolated corner, the plot of $<S_x>$ for the single slit shows that there is a small region (<100 nm) around the slit entrance where power is flowing towards the slit. There is also a larger surrounding region (~100 nm – 200 nm) where the power is flowing away from the slit entrance. In other words, SPs are generated at and propagate away from the corners of the slit entrance in the same way they are generated at an isolated corner, but there is a localized funneling effect towards the slit. If light from outside the slit region is funneled into the slit, this increases the effective size of the slit region (in terms of light-capturing ability, similar to the idea of an effective cross section in scattering theory). Note that this region is quite small for the single slit case, so the effective size of the slit is not increased substantially.

For multiple slits, the $<S_x>$ fields exhibit a quadrupolar pattern around the metallic "islands" formed by the slits. Physically this indicates that virtually all of the light falling on the slit array itself is funneled towards the entrance. This is further illustrated by the nearly uniform forward propagation of power in the slit array region shown by the $<S_y>$ plots. Outside this region, the propagation of power in the y-direction is weaker and not uniformly positive,

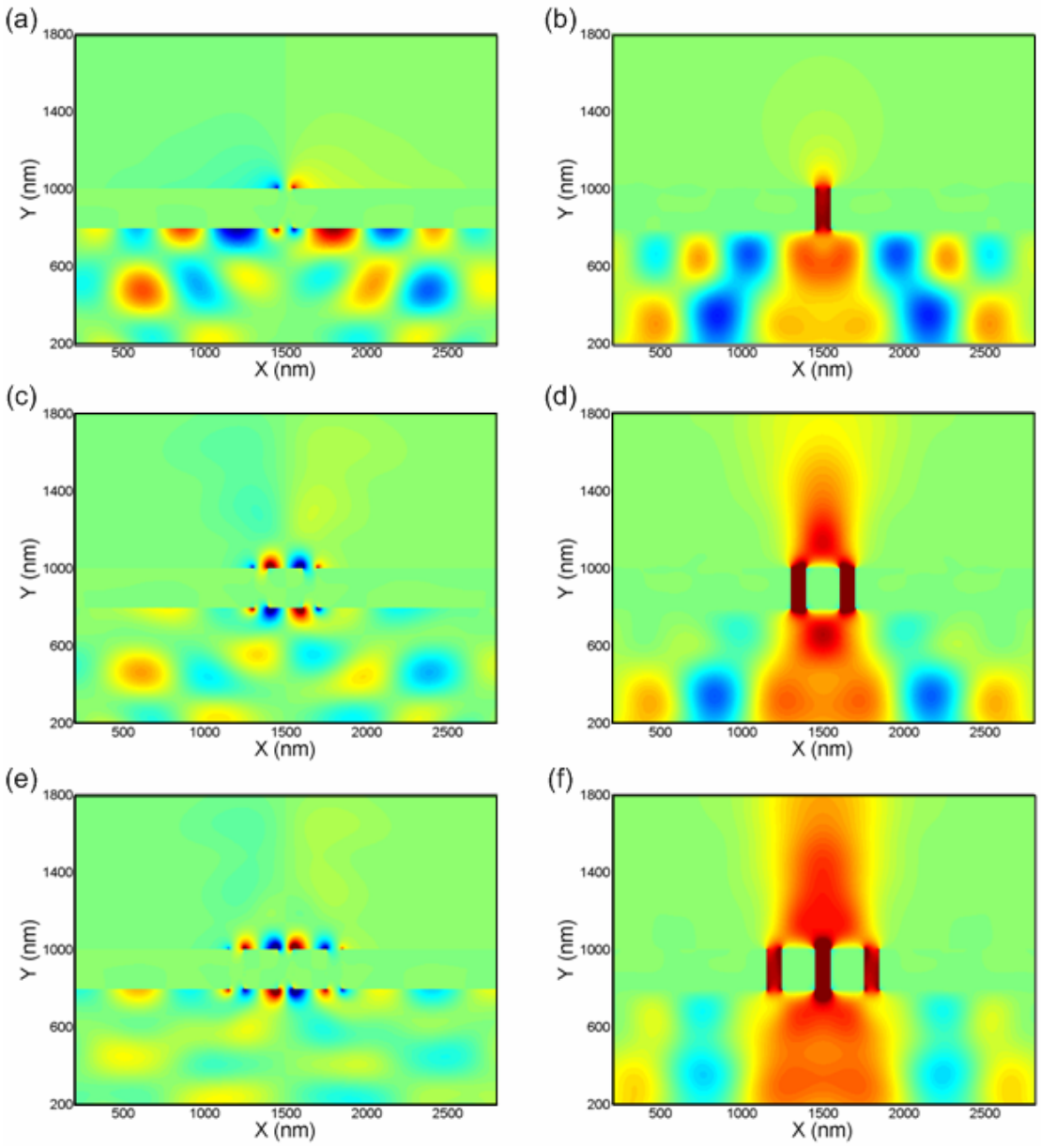

Figure 6: Time-averaged Poynting vector plots for a planewave (λ_0 = 650 nm) incident on a 200 nm thick silver film with one (a,b), two (c,d), and three (e,f) slits, each 100 nm wide; in the latter two cases, there is a center-to-center slit spacing of 300 nm. Plots of the x-component are on the left (a,c,e), plots of the y-component are on the right (b,d,f).

indicating that a significant component of the power is reflected. In the slit region, there is strong transmission and very weak reflection.

One of the most puzzling aspects of this problem is the idea that power flows towards the corner and the slit. Is the system really like a funnel in the sense that light falling outside the slit region is drawn towards it? SPs cannot be generated unless there is some periodicity or discontinuity to provide extra momentum; for the case of the isolated wedge, this would seem to imply that they can only be generated at the corner itself. If this is the case, where is this inward traveling power coming from? In the multiple slit structures, there may be significant SP-mediated power transfer between neighboring slits; but the fact that there is noticeable funneling observed in the corner and single slit structures indicates that there is more to the problem.

An important clue may lie in the fact that – at any surface – the normal component of the Poynting vector is continuous and the tangential component of the Poynting vector is discontinuous by a factor of ε_M. At the corner in Figure 5, SPs are most strongly created on the interface defined by the x = 1500 nm line. Most of the power associated with these SPs propagates in the y-direction. While it is difficult to discern in this figure, it can be shown that for a system with loss ($\mathrm{Im}(\varepsilon_M) \neq 0$), there will be some power flowing from the air side into the metal. So at this interface, there will be a small amount of power flowing in the negative x-direction, with a penetration depth on the metal side of about 20 nm (depending on wavelength). Near the corner, there may be significant coupling between the right (x = 1500 nm) and front (y $\approx$ 800 nm) faces. In other words, the power flowing into the right interface ends up flowing tangential to the front interface. While this Poynting vector is relatively small (if the loss is small), the small tangential component of the Poynting vector on the metal side is enhanced by a factor of ε_M on the air-side (which can be very large, in this case it is around -17). In summary, a

little loss on the right-interface (Poynting vector in –x-direction on the metal side), induces a significant amount of inward-directed power on the air-side of the front-interface (the negative sign of ε_M causes the change to the +x-direction). This can be simplistically viewed as power circulating around the corner. This picture makes sense, because the funneling region is approximately the same size as the penetration depth in the metal. The slight increase in the size of the funneling region in the context of the single slit also makes sense; decreasing the slit width causes the penetration depth in the metal to increase.

There are some subtle aspects to this model which make it difficult to analyze. It is unclear whether this circulation of power around the corner would account for additional funneling, or would simply recycle some of the power lost in the slit region. In order to calculate the actual size of the funneling region predicted by this model, one would need to know how much power is propagating away from the corner on both the front and right-hand interfaces.

The radiation pattern of power coming from a single slit is also a topic of interest which can be analyzed with FDTD. It can be seen from Figure 7 that the transmitted light is characterized by approximately circular wavefronts. For the single slit structure, the radial component of the time-averaged Poynting vector is shown in Figure 8 at several different distances from the slit (the center is defined as the center of the slit on the back side of the film and the angle is defined with respect to the x-axis, a sketch is shown in Figure 8(a)). It appears that the radiation pattern is relatively uniform near 90°, away from the metal surfaces. Moreover, this component appears to die off with a 1/R dependence. For a 2D system, the component of the Poynting vector that decreases as 1/R will appear in the far-field (the total power – equal to the Poynting vector times an area proportional to R – will be constant). As expected, there is also a significant component of the power which is strongly localized near the metal interface. This

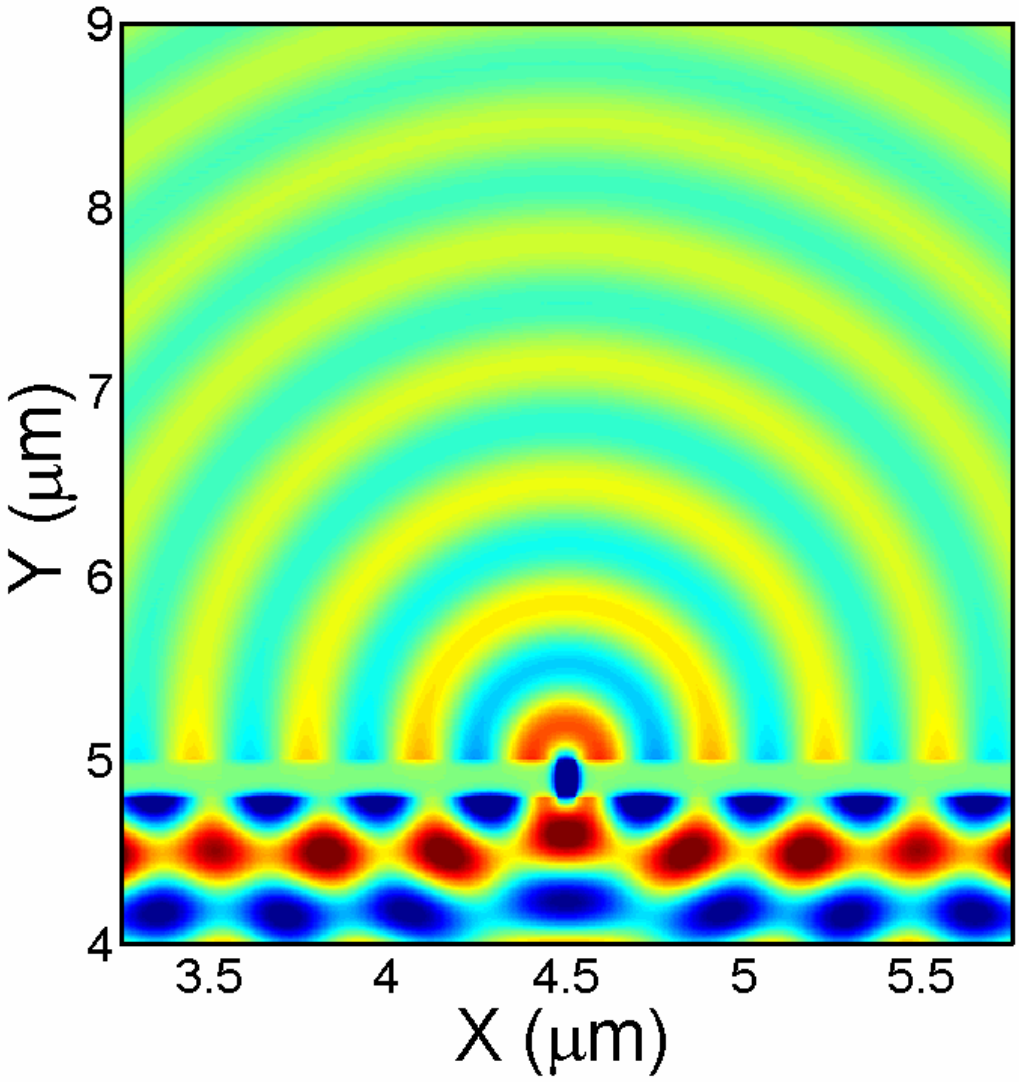

Figure 7: Magnetic field (H_z) corresponding to the light emitted from a single slit (light normally incident from bottom of problem space, λ_0 = 650 nm, 100 nm slit width, 200 nm silver thickness).

is associated with SPs and will drop off exponentially with distance. Though the propagation length of SPs is relatively large for a low-loss system (about 25 μm for a single interface at λ_0 = 650 nm), the total power associated with the SPs will go to zero in the true far-field.

Figures 9 and 10 illustrate how the power flow through a single slit changes as a function of wavelength (at λ_0 = 400 nm and 800 nm, respectively). It can be seen that the ratio of power coupled into SP modes (the peaks near the metal) to the freely transmitted power increases as wavelength decreases from ~1.7 at λ_0 = 800 nm to ~3 at λ_0 = 400 nm. It also appears from the plot of $<S_x>$ in each figure that the inward funneling region grows with wavelength, despite the fact that the metal penetration depth decreases as wavelength increases. This does not contradict the explanation for the funneling effect given above. The penetration depth decreases because $|\varepsilon_M|$ increases rapidly with wavelength above the plasmon resonance; the ratio of the tangential Poynting vector on the air side to the tangential Poynting vector on the metal side is proportional to the metallic dielectric constant, this enhances the feedback mechanism between the normal component of the Poynting vector in the slit region and the tangential component near the slit entrance.

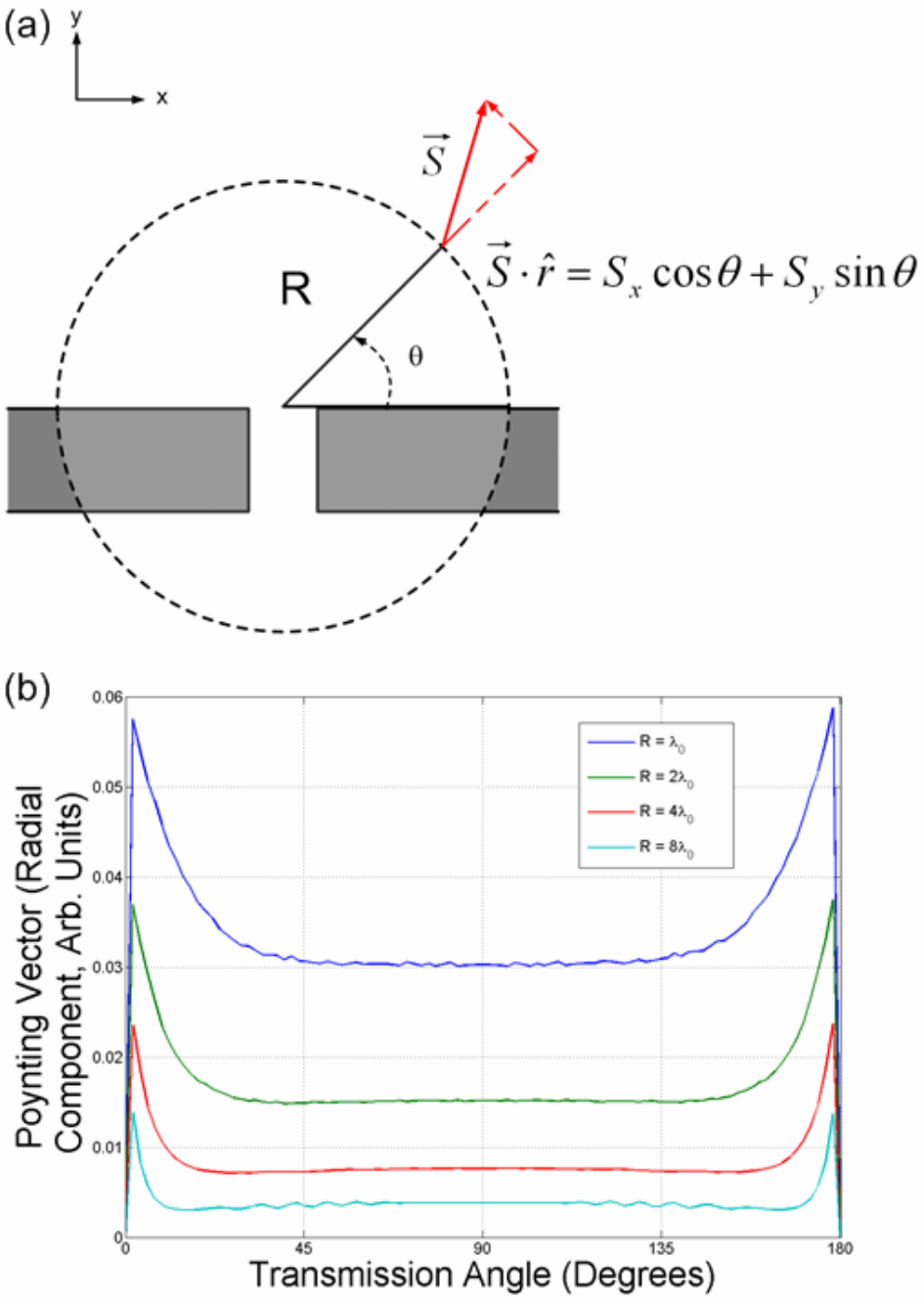

Figure 8: (a) Sketch of radial component of the Poynting vector, (b) plot of time-averaged radial Poynting vector as a function of direction with respect to the center of the slit, shows gradual transition from near-field to far-field (λ_0 = 650 nm, 100 nm wide single slit, 200 nm thick film).

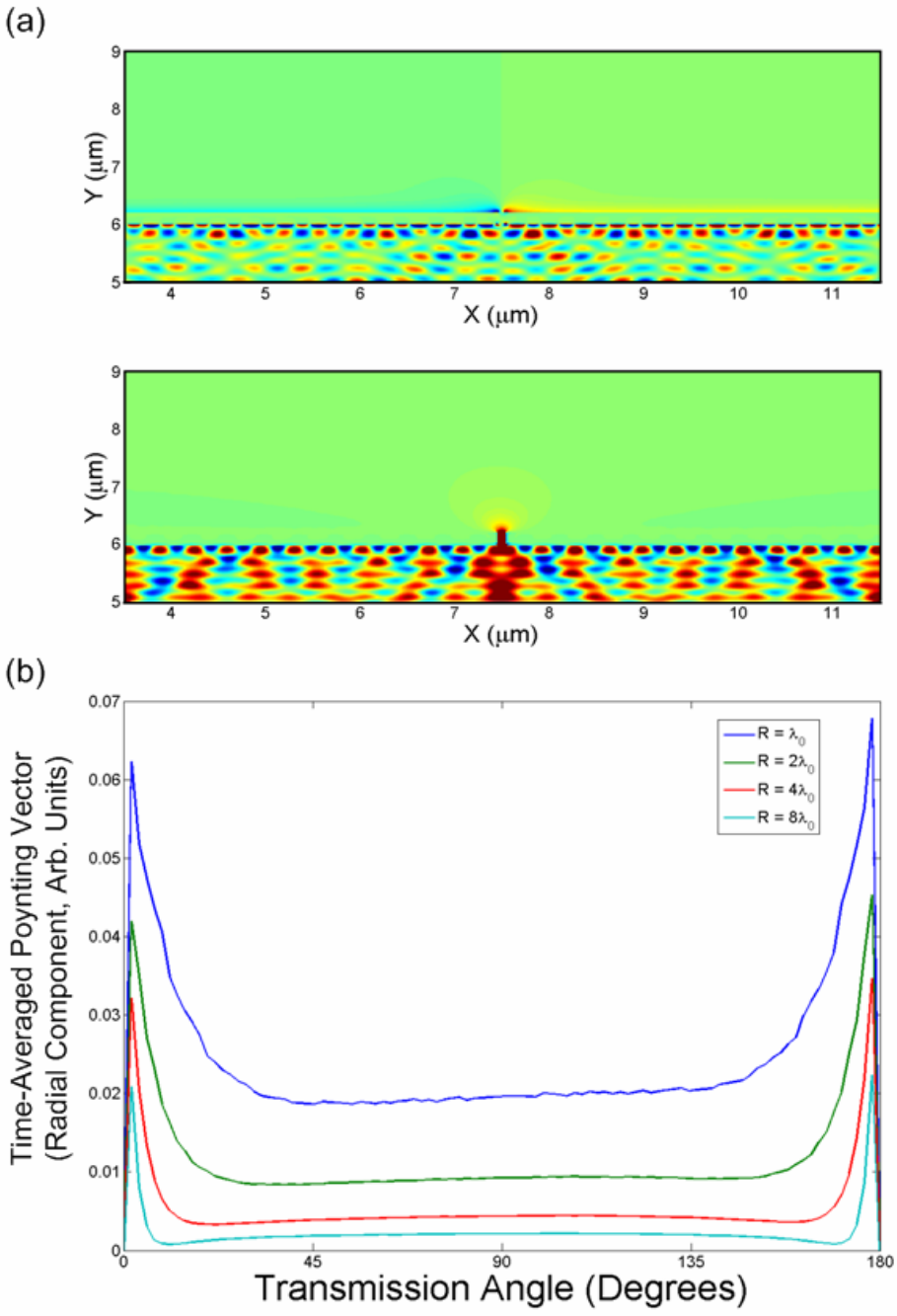

Figure 9: (a) Plot of time-averaged Poynting vector for single slit (100 nm width), x and y components (top and bottom, respectively), (b) radial component of the time-averaged Poynting vector as a function of transmission direction and distance (λ_0 = 400 nm, 100 nm wide single slit, 200 nm thick film).

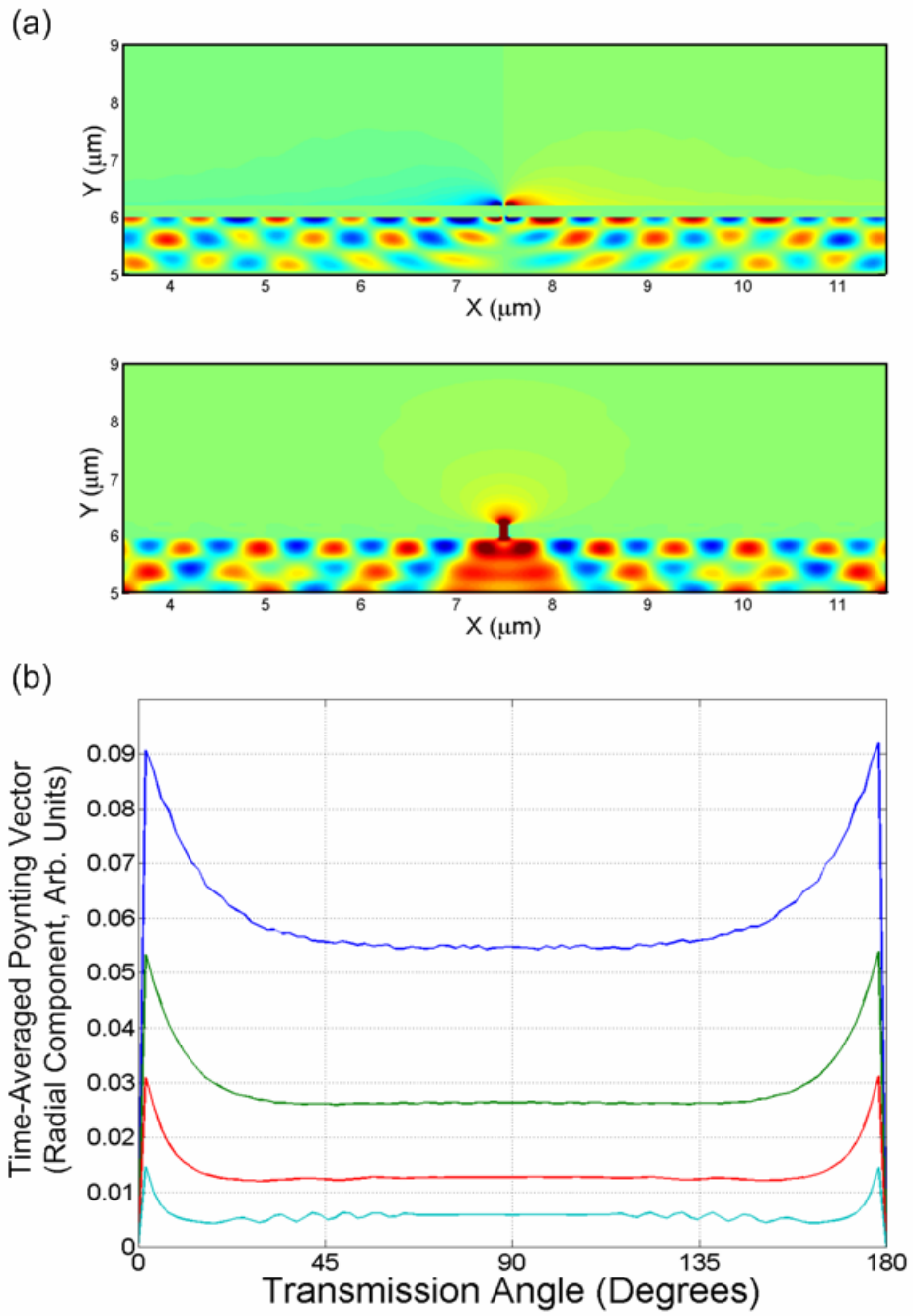

Figure 10: (a) Plot of time-averaged Poynting vector for single slit (100 nm width), x and y components (top and bottom, respectively), (b) radial component of the time-averaged Poynting vector as a function of transmission direction and distance (λ_0 = 800 nm, 100 nm wide single slit, 200 nm thick film)

3.0 CHARGE AND POWER FLOW IN PLASMONIC MODES

3.1 POLARIZATION CHARGE

A SP wave can be compared to a wave on the surface of a body of water. While there are certain areas where the water is higher (crests) and certain areas where the water is lower (troughs), the total amount of water in the ocean is unchanged by the presence of the wave. The same is true of surface plasmons. At equilibrium, the negative charge of the electron sea is cancelled out everywhere by the positive background of the nuclei in the metallic lattice. But if you imagine a ripple on the surface of the electron sea there will be crests – negatively charged regions where electrons have accumulated – and troughs – positively charged regions where electrons have depleted – but overall the system is still charge neutral. Because the system is charge neutral, we assume there is no bulk free charge.

The polarization charge on an interface can be calculated by determining an analytic expression for the discontinuity in the normal component of the electric field. Gauss' law in terms of the displacement field in a medium without free-charge is as follows.

$$\nabla \cdot \vec{D} = \nabla \cdot (\varepsilon_0 \vec{E} + \vec{P}) = 0 \tag{3.1}$$

While there is no free charge in the system, the surface plasmon wave is characterized by a propagating polarization charge disturbance at the interface between the air and metal.

$$\nabla \cdot \vec{E} = -\frac{\nabla \cdot \vec{P}}{\varepsilon_0} \equiv \frac{\rho_P}{\varepsilon_0} \tag{3.2}$$

Where ρ_P is the polarization charge density caused by the discontinuity of the electric field at the interface. The polarization charge density can be calculated from the field solutions in the previous section. Consider the interface at x = 0 (as shown in Figure 2). The integral form of Gauss' law can be applied to an infinitely small "pillbox" (Gaussian surface) with one surface on the negative side of the interface and one on the positivie side. The surface can be made infinitesimally thin enough so that the contribution of the surface integral along the sides of the pillbox will be zero.

$$\oiint \vec{E} \cdot d\vec{A} = \frac{1}{\varepsilon_0} \iiint \rho \, dV \tag{3.3}$$

$$\lim_{x \to 0} (\vec{E}^{(2)}(x,y) \cdot \hat{x} + \vec{E}^{(1)}(x,y) \cdot (-\hat{x}))dA = \frac{dQ_P}{\varepsilon_0} \tag{3.4}$$

$$\sigma_{x=0}(y) = dQ_P / dA = \lim_{x \to 0} \varepsilon_0 (E_x^{(2)}(x,y) - E_x^{(1)}(x,y)) \tag{3.5}$$

The polarization charge on the other interface (at x = a) should be similar due to symmetry. It can be derived in a similar manner.

$$\sigma_{x=a}(y) = \lim_{x \to a} \varepsilon_0 (E_x^{(3)}(x,y) - E_x^{(2)}(x,y)) \tag{3.6}$$

For the symmetric mode, the x-component of the electric field should be symmetric with respect to the center of the slit. For the anti-symmetric mode, the x-component of the electric field

should be anti-symmetric with respect to the center of the slit (x = a/2) (region 2). Likewise, the fields on the metal side of the boundaries can be related through symmetry.

$$E_x^{(1)}(0,y) = \pm E_x^{(3)}(a,y) \tag{3.7}$$

Putting the above equations together illustrates the symmetry of the two polarization charge waves.

$$\sigma_{x=0}^{\pm}(y) = \lim_{x\to 0} \varepsilon_0 (E_x^{(2)}(x,y) - E_x^{(1)}(x,y)) \tag{3.8}$$

$$\sigma_{x=0}^{\pm}(y) = \lim_{x\to a} \varepsilon_0 (\pm E_x^{(2)}(x,y) \mp E_x^{(3)}(x,y)) \tag{3.9}$$

$$\sigma_{x=0}^{\pm}(y) = \mp \lim_{x\to a} \varepsilon_0 (E_x^{(3)}(x,y) - E_x^{(2)}(x,y)) \tag{3.10}$$

$$\sigma_{x=0}^{\pm}(y) = \mp \sigma_{x=a}^{\pm}(y) \tag{3.11}$$

This shows that the polarization surface charge density associated with the symmetric mode will be anti-symmetric and the charge density for the anti-symmetric mode will be symmetric. This could cause confusion, so it helpful to remember that the symmetry of the mode has (arbitrarily) been defined so that it refers to the symmetry of the H_z field. Though the information provided by the symmetry is enlightening, a full analytic expression for the polarization charge density can be calculated by inserting the field expressions from 2.25-27 into 3.5 or 3.6.

$$\sigma_{x=0}^{\pm}(y) = \lim_{x\to 0} \varepsilon_0 \eta_0 H_0 \left(\frac{\beta}{k_0}\right)\left[-\frac{1}{\varepsilon_D}(e^{-\gamma_D x} \pm e^{-\gamma_D (a-x)}) + \frac{1}{\varepsilon_M}(1 \pm e^{-\gamma_D a}) e^{\gamma_M x}\right] e^{i\beta y} \tag{3.12}$$

$$\sigma_{x=0}^{\pm}(y) = \frac{H_0}{c}\left(\frac{\beta}{k_0}\right)(1 \pm e^{-\gamma_D a})\left(\frac{\varepsilon_D - \varepsilon_M}{\varepsilon_M \varepsilon_D}\right) e^{i\beta y} \tag{3.13}$$

And the polarization charge at x = a follows from 3.11.

$$\sigma^{\pm}_{x=a}(y) = \mp \frac{H_0}{c}\left(\frac{\beta}{k_0}\right)(1 \pm e^{-\gamma_D a})\left(\frac{\varepsilon_D - \varepsilon_M}{\varepsilon_M \varepsilon_D}\right)e^{i\beta y} \tag{3.14}$$

Along the propagation direction, the polarization charge will oscillate with the same period as the fields (they both vary sinusoidally as the real part of $e^{i\beta y}$). The amplitude of the charge wave has four major components which are multiplied together: a normalization constant in front to give the proper units of charge per area $\left(\frac{H_0}{c}\right)$, an expression related to the dielectric contrast between the two media $\left(\frac{\varepsilon_D - \varepsilon_M}{\varepsilon_M \varepsilon_D}\right)$, the effective index $\left(N \equiv \frac{\beta}{k_0}\right)$, and an additional coupling factor $\left(1 \pm e^{-\gamma_D a}\right)$. The dielectric contrast is entirely dependent upon the material properties of the system, while the effective index and coupling factor are dependent upon both the material and the geometry (the slit width). If both the effective index and dielectric constants are real, the polarization charge will be in phase with the magnetic field and 90º out of phase with the y-component of the electric field. If there is loss in the system, there will be a phase shift between the polarization charge and the magnetic field.

3.2 DYNAMIC AND TIME-AVERAGED POWER FLOW

The evolution of the surface charge distribution is connected to the energy flow across the interface. One way of looking at this is by deriving the Poynting vector. The Poynting vector is a measure of the power flux density associated with an electromagnetic field; a proof of this

relationship and the Poynting vector's properties are discussed in most electromagnetic theory texts [3].

$$\vec{\mathbf{S}}(x,y,t) = \vec{\mathbf{E}}(x,y,t) \times \vec{\mathbf{H}}(x,y,t) \tag{3.15}$$

It should be pointed out that the fields in 3.15 are real, time-dependent functions derived by taking the real or imaginary part of the fields derived in complex notation above. The time-averaged Poynting vector, on the other hand, is most easily described in complex notation.

$$\left\langle \vec{\mathbf{S}} \right\rangle (x,y) = \frac{1}{2} \mathrm{Re}\left[\vec{\mathbf{E}}(x,y) \times \vec{\mathbf{H}}^{*}(x,y) \right] \tag{3.16}$$

The time period of optical light is roughly on the order of femtoseconds (1 fs =10^{-15} s), which is much shorter than most physical measurements. As a result, most important physical properties of the optical field can be analyzed with the time-averaged Poynting vector alone. The time-dependent Poynting vector serves to provide some intuition about how power flows dynamically in a system.

Some interesting features associated with the Poynting vector field can be qualitatively deduced from the general expression in 3.16. The boundary conditions require that the tangential components of both the magnetic and electric fields (H_z and E_y) should be continuous across the interface; this forces the normal component of the time-averaged Poynting vector ($<S_x>$) to also be continuous. The continuity of the normal component of the Poynting vector across an interface is equivalent to conservation of power flow across that surface. In other words, the surface itself does not absorb or generate any power. The continuity of normal component of the electric flux density (D_x) requires that the normal component of the electric field (E_x) be

discontinuous by a factor of $\varepsilon_D/\varepsilon_M$. If Re($\varepsilon_M$) is assumed to be negative, E_x will switch signs across the interface and the longitudinal component of the Poynting vector (S_y) will also be pointed in the opposite direction on the metal side. So if power in the slit region is flowing in the +y direction, the power on the metal side of the slit walls will be flowing in the –y direction. The reversal of energy flow has been analyzed in idealized, lossless metal-dielectric systems [16],[20],[21]. This property has also been observed in systems with negative index materials (where both the permittivity and permeability can be negative) [22],[17].

Like the fields, the Poynting vector is described in a piece-wise fashion in each of the three regions – the dielectric region inside the slit and the metal regions outside the slit.

$$< S_y >^{(1)} (x,y) = \frac{|H_0|^2}{2}\eta_0 \,|1 \pm \exp(-\gamma_D a)|^2 \,\mathrm{Re}\left(\frac{N_{eff}}{\varepsilon_M}\right) \times \exp[2\,\mathrm{Re}(\gamma_M)x - 2\,\mathrm{Im}(\beta)y] \tag{3.17}$$

$$< S_y >^{(2)} (x,y) = |H_0|^2 \,\eta_0 \,\mathrm{Re}\left(\frac{N_{eff}}{\varepsilon_D}\right)\exp[-\mathrm{Re}(\gamma_D)a - 2\,\mathrm{Im}(\beta)y] \times \{\cosh[\mathrm{Re}(\gamma_D)(a-2x)] \pm \cos[\mathrm{Im}(\gamma_D)(a-2x)]\} \tag{3.18}$$

$$< S_y >^{(3)} (x,y) = \frac{|H_0|^2}{2}\eta_0 \,|1 \pm \exp(-\gamma_D a)|^2 \,\mathrm{Re}\left(\frac{N_{eff}}{\varepsilon_M}\right) \times \exp[2\,\mathrm{Re}(\gamma_M)(a-x) - 2\,\mathrm{Im}(\beta)y] \tag{3.19}$$

$$< S_x >^{(1)} (x,y) = \frac{|H_0|^2}{2}\eta_0 \,|1 \pm \exp(-\gamma_D a)|^2 \,\mathrm{Im}\left(\frac{\gamma_M}{k_0\varepsilon_M}\right) \times \exp[2\,\mathrm{Re}(\gamma_M)x - 2\,\mathrm{Im}(\beta)y] \tag{3.20}$$

$$< S_x >^{(2)} (x,y) = |H_0|^2 \,\eta_0 \exp[-\mathrm{Re}(\gamma_D)a - 2\,\mathrm{Im}(\beta)y] \times \left\{\mathrm{Re}\left(\frac{\gamma_D}{\varepsilon_D k_0}\right)\sin[\mathrm{Im}(\gamma_D)(2x-a)] \mp \mathrm{Im}\left(\frac{\gamma_D}{\varepsilon_D k_0}\right)\sinh[\mathrm{Re}(\gamma_D)(a-2x)]\right\} \tag{3.21}$$

$$< S_x >^{(3)} (x,y) = -\frac{|H_0|^2}{2}\eta_0 \,|1 \pm \exp(-\gamma_D a)|^2 \,\mathrm{Im}\left(\frac{\gamma_M}{k_0\varepsilon_M}\right) \times \exp[2\,\mathrm{Re}(\gamma_M)(a-x) - 2\,\mathrm{Im}(\beta)y] \tag{3.22}$$

If a lossless dielectric function is assumed for the metal in a system that supports SPs ($\mathrm{Im}(\varepsilon_M) = 0$), then $\mathrm{Im}(\beta) = \mathrm{Im}(\gamma_D) = \mathrm{Im}(\gamma_M) = 0$. It is then obvious from 3.20-3.22 that for a lossless system the transverse component (*x*-component) of the time-averaged Poynting vectors would also be zero everywhere and there would be no net power transfer across the interface over a given period. It is also confirmed from 3.17-3.19 that there would be no energy loss along the propagation direction (the y-component of the Poynting vector would be constant with respect to y). Metals, however, usually exhibit a complex dielectric constant; as a result, all three propagation/decay constants (k_{sp}, γ_D, and γ_M) will also be complex. In this case, the x-component of the time-averaged Poynting vector would be non-zero and would point towards the metal side in all regions. As it propagates, energy would be lost due to the non-zero conductivity of the metal. The energy flow associated with the transverse component of the Poynting vector on the dielectric side supplies the energy that is absorbed in the metal region. It is important to note that the energy loss occurs only in the metal side, and not in the dielectric side. This can be proven by calculating $\vec{\nabla}\cdot\vec{S}$ in each region. The slit region is usually composed of some lossless dielectric – e.g., air or water.

$$\partial_y \langle S_y \rangle^{(2)} = -2\,\mathrm{Im}(k_{sp}) \langle S_y \rangle^{(2)} \tag{3.23}$$

$$\begin{aligned} \partial_x \langle S_x \rangle^{(2)} &= 2\eta_0 \,|\, H_0 \,|^2 \, e^{-\mathrm{Re}(\gamma_D)a - 2\,\mathrm{Im}(k_{sp})y} \\ &\times \left\{ \mathrm{Re}(\gamma_D)\,\mathrm{Im}\left(\frac{\gamma_D}{\varepsilon_D k_0}\right) Cosh[\mathrm{Re}(\gamma_D)(2x-a)] \pm \mathrm{Im}(\gamma_D)\,\mathrm{Re}\left(\frac{\gamma_D}{\varepsilon_D k_0}\right) Cos[\mathrm{Im}(\gamma_D)(2x-a)] \right\} \end{aligned} \tag{3.24}$$

The free-space wavevector k_0 is proportional to the inverse wavelength and is a real number. The material in the slit region is assumed to be lossless, meaning the dielectric constant is a real number. This enables simplification of the above equations.

$$\partial_x \langle S_x \rangle^{(2)} = \frac{2\eta_0 \mid H_0 \mid^2}{\varepsilon_D k_0} e^{-\mathrm{Re}(\gamma_D)a - 2\,\mathrm{Im}(k_{sp})y}\, \mathrm{Re}(\gamma_D)\,\mathrm{Im}(\gamma_D) \times \{Cosh[\mathrm{Re}(\gamma_D)(2x-a)] \pm Cos[\mathrm{Im}(\gamma_D)(2x-a)]\} \tag{3.25}$$

$$\partial_y \langle S_y \rangle^{(2)} = -2\frac{\eta_0 \mid H_0 \mid^2}{k_0 \varepsilon_D} \mathrm{Re}(k_{sp})\mathrm{Im}(k_{sp}) e^{-\mathrm{Re}(\gamma_D)a - 2\,\mathrm{Im}(k_{sp})y} \times \{Cosh[\mathrm{Re}(\gamma_D)(2x-a)] \pm Cos[\mathrm{Im}(\gamma_D)(2x-a)\} \tag{3.26}$$

The divergence of the Poynting vector is simply the sum of 3.25 and 3.26.

$$\nabla \cdot \langle \mathbf{S} \rangle^{(2)} = 2\frac{\eta_0 \mid H_0 \mid^2}{k_0 \varepsilon_D} \left[\mathrm{Re}(\gamma_D)\,\mathrm{Im}(\gamma_D) - \mathrm{Re}(k_{sp})\mathrm{Im}(k_{sp})\right] e^{-\mathrm{Re}(\gamma_D)a - 2\,\mathrm{Im}(k_{sp})y} \times \{Cosh[\mathrm{Re}(\gamma_D)(2x-a)] \pm Cos[\mathrm{Im}(\gamma_D)(2x-a)\} \tag{3.27}$$

Consider the relationship between the plasmon wavevector and γ_D.

$$\gamma_D^2 = k_{sp}^2 - \varepsilon_D k_0^2 \tag{3.28}$$

$$[\mathrm{Re}(\gamma_D) + i\,\mathrm{Im}(\gamma_D)]^2 = [\mathrm{Re}(k_{sp}) + i\,\mathrm{Im}(k_{sp})]^2 - \varepsilon_D k_0^2 \tag{3.29}$$

$$[\mathrm{Re}(\gamma_D)^2 - \mathrm{Im}(\gamma_D)^2] + 2i\,\mathrm{Re}(\gamma_D)\,\mathrm{Im}(\gamma_D) = [\mathrm{Re}(k_{sp})^2 - \mathrm{Im}(k_{sp})^2] + 2i\,\mathrm{Re}(k_{sp})\,\mathrm{Im}(k_{sp}) - \varepsilon_D k_0^2 \tag{3.30}$$

The $\varepsilon_D k_0^2$ term is real, so by taking the imaginary part of the left and right sides, only two of the five terms remain.

$$\mathrm{Re}(\gamma_D)\,\mathrm{Im}(\gamma_D) = \mathrm{Re}(k_{sp})\,\mathrm{Im}(k_{sp}) \tag{3.31}$$

Finally, 3.31 can be used with 3.27 to show that the divergence in the slit region is zero.

$$\nabla \cdot \langle \mathbf{S} \rangle^{(2)} = 0 \tag{3.32}$$

Physically this means that the total time-averaged power entering / leaving any infinitesimal region on the dielectric side is equal to zero. In other words, there is no loss or gain anywhere in the slit region – which is to be expected. It can be shown in a similar manner that $\nabla \cdot S > 0$ on the metal side; meaning that in any infinitesimal region on the metal side there is more power entering than leaving, indicating loss.

The total power carried in each region (the dielectric gap and the left and right metal regions) can be calculated by integrating 3.17-3.19 over the xz-plane (for $x < 0$, $0 < x < a$, and $x > a$, respectively, and over an arbitrary distance along the z-direction).

$$\left(\frac{P}{\delta z}\right)^{(1)} = \left(\frac{P}{\delta z}\right)^{(3)} = \frac{|H_0|^2}{4}\eta_0 \frac{|1 \pm \exp(-\gamma_D a)|^2}{\mathrm{Re}(\gamma_M)} \mathrm{Re}\left(\frac{N_{eff}}{\varepsilon_M}\right) \exp\left[-2\,\mathrm{Im}(k_{sp})y\right] \tag{3.33}$$

$$\begin{aligned}\left(\frac{P}{\delta z}\right)^{(2)} &= |H_0|^2 \eta_0 \,\mathrm{Re}\left(\frac{N_{eff}}{\varepsilon_D}\right) \exp[-2\,\mathrm{Im}(k_{sp})y - \mathrm{Re}(\gamma_D)a] \\ &\times \left\{\frac{\sinh[\mathrm{Re}(\gamma_D)a]}{\mathrm{Re}(\gamma_D)} \pm \frac{\sin[\mathrm{Im}(\gamma_D)a]}{\mathrm{Im}(\gamma_D)}\right\}\end{aligned} \tag{3.34}$$

These results will be discussed in depth in subsequent sections. The longitudinal component of the Poynting vector and the power flow in the direction of propagation will both be symmetric with respect to the center of the slit, regardless of whether the fields are symmetric or anti-symmetric.

3.3 SINGLE INTERFACE

A single metal-dielectric planar interface is inarguably the simplest plasmonic system imaginable. Before building up to the slightly more complicated, but more physically interesting

single slit structure, the power and charge results will be applied to this system. This will help isolate the purely material effects of changing wavelength from the effect of coupling caused by having to slit walls interfering with each other.

3.3.1 As Limiting Case of General Result

The physics of a single metallic interface can be analyzed as a special case of the single infinite slit in the limit as the slit width goes to infinity. More accurately, the two sides of the slit can be accurately modeled as two uncoupled metal-dielectric interfaces when the separation is much larger than the penetration depth of the fields in the air region ($1/\gamma_{D,}$ usually on the order of a few hundred nanometers). The waveguide like modes which were discussed in the previous section will be ignored, because in this limit they will turn into unbound propagating modes, essentially reflecting off the metal surface in the dielectric half-space – not particularly interesting in this context. The new eignevalue equation can be calculated by utilizing the following limit of the hyperbolic tangent in 2.28.

$$\lim_{x \to \infty} \tanh^{\pm 1}(x) = \lim_{x \to \infty} \left(\frac{e^{x} + e^{-x}}{e^{x} - e^{-x}} \right) = 1 \tag{3.35}$$

$$\frac{\gamma_D}{\varepsilon_D} = -\frac{\gamma_M}{\varepsilon_M} \tag{3.36}$$

This result is consistent with the solution shown in 2.16 – 2.18. Now it is a relatively simple matter to apply the same limit to the expressions for the time-average Poynting vector and the polarization charge derived in Sections 3.1 and 3.2.

$$< S_y >^{(1)} (x,y) = \frac{|H_0|^2}{2}\eta_0 \,\mathrm{Re}\left(\frac{N_{eff}}{\varepsilon_M}\right)\exp[2\,\mathrm{Re}(\gamma_M)x - 2\,\mathrm{Im}(\beta)y] \tag{3.37}$$

$$< S_y >^{(2)} (x,y) = \frac{|H_0|^2}{2}\eta_0 \,\mathrm{Re}\left(\frac{N_{eff}}{\varepsilon_D}\right)\exp[-2\,\mathrm{Re}(\gamma_D)x - 2\,\mathrm{Im}(\beta)y] \tag{3.38}$$

$$< S_x >^{(1)} (x,y) = \frac{|H_0|^2}{2}\eta_0 \,\mathrm{Im}\left(\frac{\gamma_M}{k_0\varepsilon_M}\right)\exp[2\,\mathrm{Re}(\gamma_M)x - 2\,\mathrm{Im}(\beta)y] \tag{3.39}$$

$$< S_x >^{(2)} (x,y) = -\frac{|H_0|^2}{2}\eta_0 \,\mathrm{Im}\left(\frac{\gamma_D}{\varepsilon_D k_0}\right)\exp[-2\,\mathrm{Re}(\gamma_D)x - 2\,\mathrm{Im}(\beta)y] \tag{3.40}$$

$$< S_x >^{(3)} (x,y) = -\frac{|H_0|^2}{2}\eta_0 \,|1+\exp(-\gamma_D a)|^2 \,\mathrm{Im}\left(\frac{\gamma_M}{k_0\varepsilon_M}\right) \times \exp[\,2\,\mathrm{Re}(\gamma_M)(a-x) - 2\,\mathrm{Im}(\beta)y] \tag{3.41}$$

$$\sigma_{x=0}(y) = \frac{H_0}{c} N_{eff}\left(\frac{\varepsilon_D - \varepsilon_M}{\varepsilon_M \varepsilon_D}\right) e^{i\beta y} \tag{3.42}$$

For the case of a single interface, the time-averaged Poynting vector has an interesting property. Its direction, given by $A\tan(\langle S_x\rangle/\langle S_y\rangle)$, is constant within each region. To remove the effect of the y-component of the Poynting vector changing sign across the interface, the angles are defined with respect to the +y direction on the dielectric side and with respect to the –y direction on the metal side.

$$\phi_M = A\tan[\mathrm{Im}(\gamma_M/\varepsilon_M)/\mathrm{Re}(\beta/\varepsilon_M)] \tag{3.43}$$

$$\phi_D = A\tan[\mathrm{Im}(\gamma_D/\varepsilon_D)/\mathrm{Re}(\beta/\varepsilon_D)] \tag{3.44}$$

It can be seen from looking at the general expressions for the time-averaged Poynting vector in the previous section that while the 3.43 is still valid for a slit structure, 3.44 is not. The

orientation of the Poynting vector in the slit region changes significantly as a function of position within the slit region due to the interference between the plasmon waves on the two interfaces.

When the magnitude of the metallic dielectric constant is extremely large (significantly below the plasmon frequency – say at 600 nm or 800 nm free-space wavelength or more), the Poynting vector is mostly parallel to the direction of propagation. As $|\varepsilon_M|$ decreases approaching resonance, the metal will become lossier and there will be more significant power flow on the metal side. This means that there is more power supplied to the metal side which is lost to absorption and there will be a significant transverse component to the average power flow. Calculations from equations 3.43 and 3.44 show that the angle of the Poynting vector goes from 0.5^o on the air side (8.4^o in the metal side) at 650 nm free-space wavelength to 13^o on the air side (26^o on the metal side) at 350 nm. Note that the angle will always be larger on the metal side if $|\varepsilon_M| > |\varepsilon_D|$; while the normal component of the Poynting vector is continuous, the tangential component on the metal side will always be smaller by a factor of $|\varepsilon_D|/|\varepsilon_M|$.

3.3.2 Effect of Wavelength

In the single-interface case, the only degrees of freedom of the system are the dielectric constants of two regions. In this system, the effect of geometry (slit width) is removed, leaving only the effect of the dielectric function and its wavelength / frequency dependence. In silver, the dielectric constant is largely negative with a relatively small imaginary part around 650 nm wavelength. The real part gradually decreases in its magnitude and crosses -1 at around 337 nm wavelength and zero at about 324 nm [5],[6].

In Figure 11, the time-dependent and time-averaged Poynting vector fields are shown for a single metal-air interface at x = 0: (a) λ_0 = 650 nm (ε_{Ag} = -17+*i*1.2), (b) λ_0 = 400 nm (ε_{Ag} = -3.8+*i*0.68), and (c) λ_0 = 350 nm (ε_{Ag} = -1.8+*i*0.60) [6].

A few interesting properties of the Poynting vector fields can be observed from Figure 11. First, both the time-averaged and time-dependent Poynting vector magnitudes attenuate along the propagation direction. The rate of attenuation increases as the wavelength decreases and dielectric constant approaches plasmon resonance. At resonance, the imaginary part of the dielectric constant – which is associated with loss– is peaked and the real part crosses zero (around 324 nm). Loss also tends to increase because power tends to shift towards the metal side as the magnitude of the metallic dielectric constant decreases.

It can also be observed that the SP wavelength decreases substantially as frequency increases, partially due to the decrease in the free-space wavelength and partially due to the steady increase in the effective index of refraction seen by the surface plasmon. Using the expression in 2.16, the real part of the index of refraction for a single silver interface (β/k_0) is 1.03 at 650 nm, 1.15 at 400 nm and 1.36 at 350 nm.

In many papers on plasmonic systems, the dielectric constant is assumed to be entirely real. It can easily be seen from 2.16 that the index of refraction should diverge as ε_M approaches -ε_D. In a real system, the interplay between the real and imaginary parts of the dielectric constant determines the maximum effective index that the system can attain. This will be discussed in depth in Chapter 4.

It is helpful to see plots of the electric and magnetic fields next to plots of the corresponding polarization charge waves and Poynting vector field. This is done over a similar

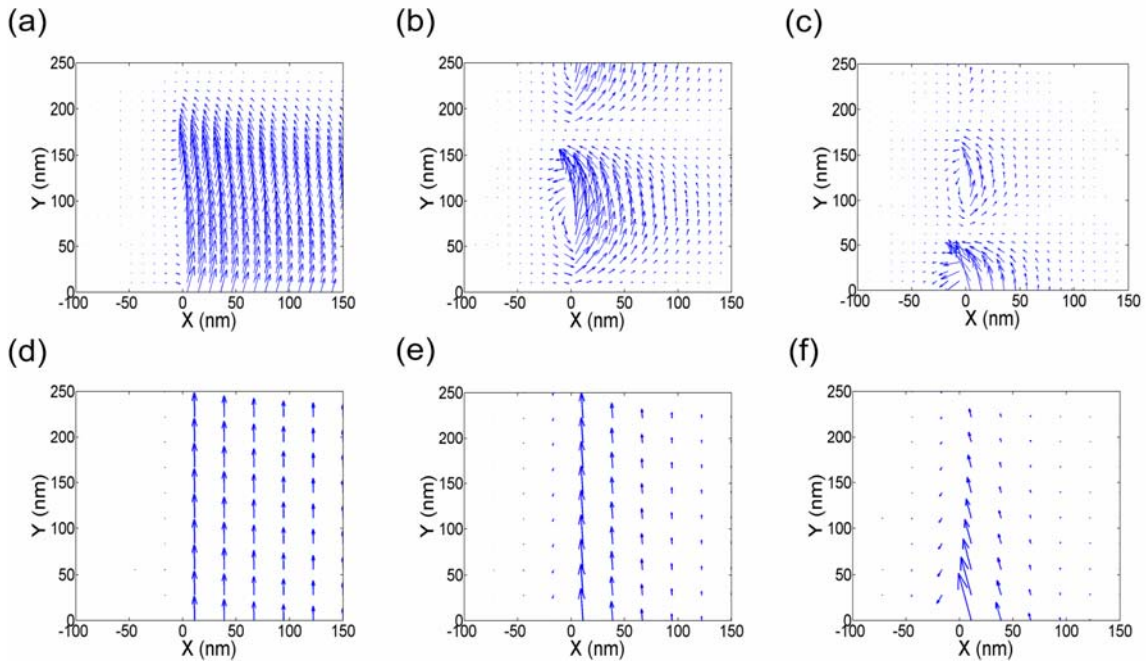

Figure 11: The time-dependent Poynting vectors (top panel) and time-averaged Poyning vectors (bottom panel) calculated for a surface plasmon wave propagating along the interface of silver ($x < 0$) and air ($x > 0$) at (a,d) 650 nm, (b,e) 400 nm, and (c,f) 350 nm free-space wavelength [23].

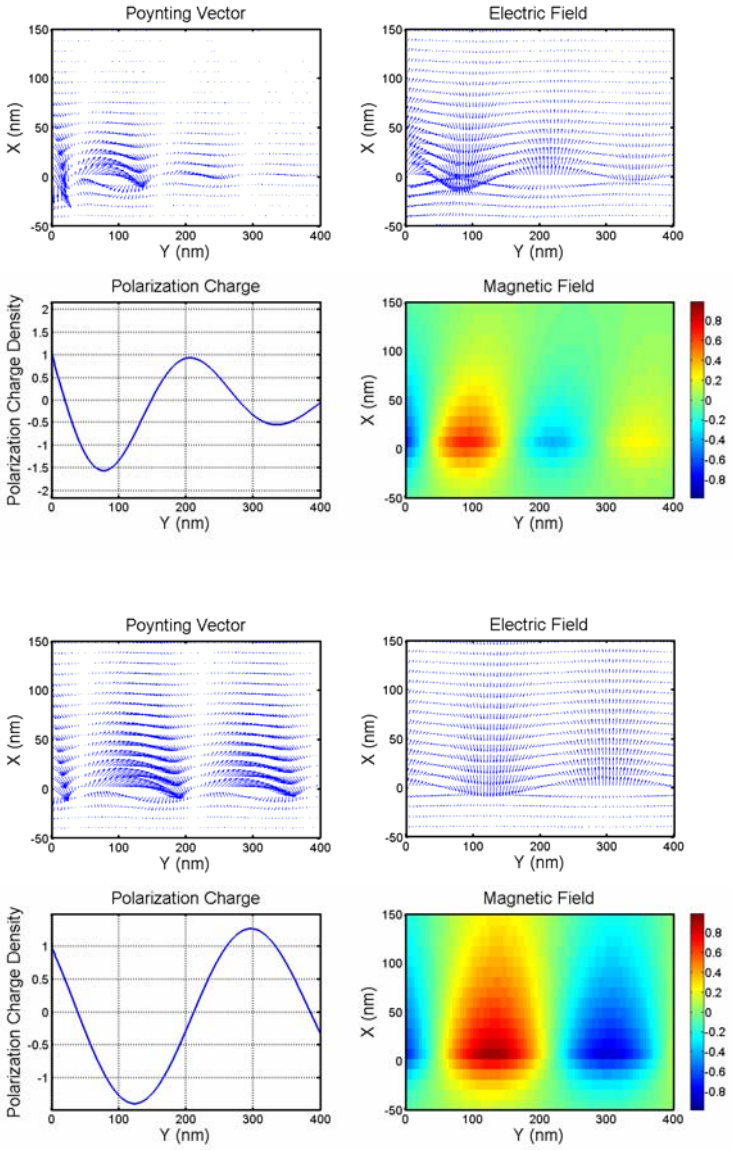

Figure 12: Instantaneous plots of the time-dependent Poynting vector, electric field, surface polarization charge and magnetic field a particular time-point at λ_0 = 350 nm (top) and λ_0 = 400 nm (bottom). The geometry is a single interface: silver fills the x <0 half-space and air fills the x>0 half-space.

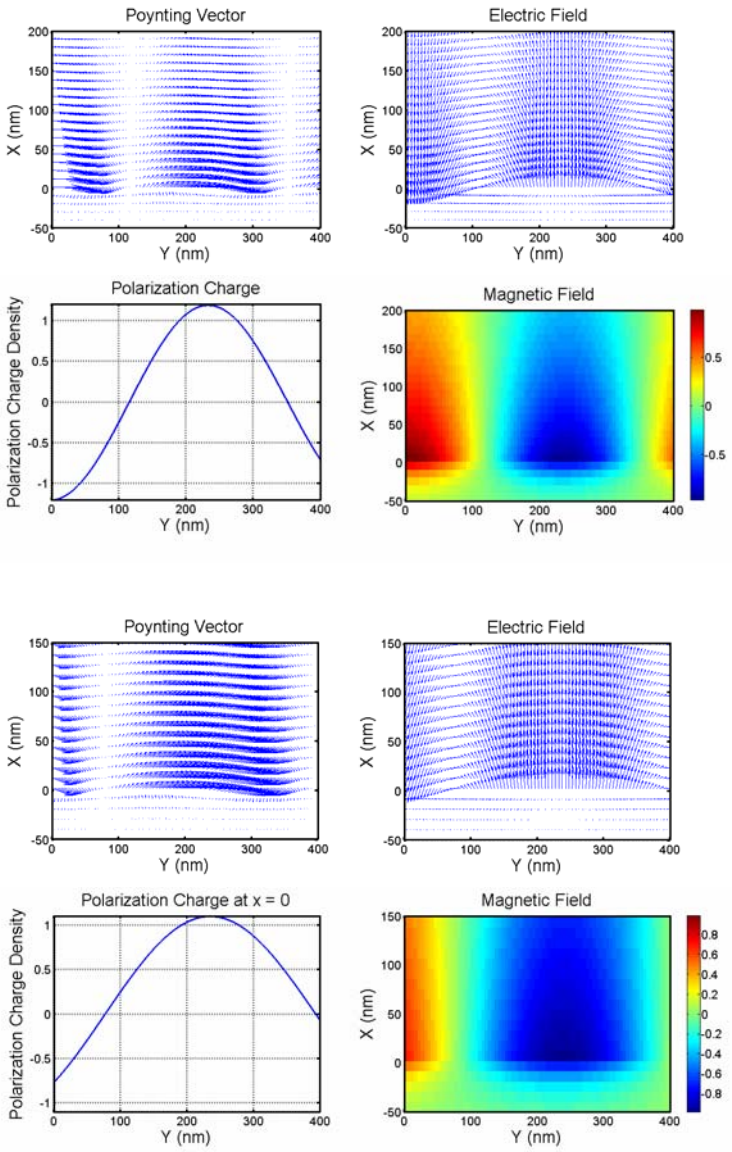

Figure 13: Instantaneous plots of the time-dependent Poynting vector, electric field, surface polarization charge and magnetic field a particular time-point at λ_0 = 500 nm (top) and λ_0 = 650 nm (bottom). Same geometry as Figure 12.

range of wavelengths in Figures 12 and 13. The interaction between the power flow, polarization charge wave and fields merits discussion – particularly in the context of the Poynting vorticity. This will be covered in Section 3.3.4, after a brief discussion of exactly what role vortex structure plays in plasmon support.

3.3.3 Regions of Plasmon Support

One interesting property of the time-dependent Poynting vector fields shown in the previous section is the manifestation of the SP as a train of vortices propagating down the interface. This circulation seems to arise from three properties of the fields. The *y*-component of the Poynting vector oscillates with twice the frequency of the fields and takes only positive values on the air side (and negative values on the metal side). In other words, the power is uniformly forward propagating on the air-side and uniformly backward propagating on the metal-side. This is because the H_z and E_x fields oscillate in phase (from 2.19-2.27, assuming β is real).

The *x*-component of the Poynting vector also oscillates with half the field wavelength (as 2βy), but oscillates between positive and negative values. This results from H_z and E_y oscillating about 90° out of phase (satisfied if γ_D and γ_M are real – implying that the wave is surface bound).

In other words, the first two results are the direct result of a surface bound wave propagating along the interface [20],[24]. The expression in 3.36 requires that there must be a sign change in the dielectric constant across the interface if the wave is surface bound. Otherwise the field would decay away from the interface in one direction and grow exponentially away from the interface in the other direction – a thoroughly unphysical result. So the reversal of power flow on the metal side is directly related to the other two requirements.

Consider a planar interface with a dielectric on the x > 0 side ($\varepsilon_1 > 0$ and $\mu_1 > 0$) and some other material, which may have positive or negative permittivity and permeability on the other side (ε_2 and μ_2 for $x < 0$). A more general system was analyzed by Tsakmakidis et al. [17], but we are primarily interested in describing whether a particular material system can or cannot support surface bound electromagnetic waves. By definition of a surface bound propagating wave, the fields must be oscillatory along the interface (*y*-direction, with effective index of refraction N_{eff}) and evanescent normal to the interface (+*x* and –*x* directions, with decay constants γ_1 and γ_2). The TE and TM modes must each satisfy three relationships.

$$\beta^2 - \gamma_1^2 = \varepsilon_1 \mu_1 k_0^2 \tag{3.45}$$

$$\beta^2 - \gamma_2^2 = \varepsilon_2 \mu_2 k_0^2 \tag{3.46}$$

$$\frac{\gamma_2}{\mu_2} = -\frac{\gamma_1}{\mu_1} \quad \text{(TE Only)} \tag{3.47}$$

$$\frac{\gamma_2}{\varepsilon_2} = -\frac{\gamma_1}{\varepsilon_1} \quad \text{(TM Only)} \tag{3.48}$$

The quantities γ_1 and γ_2 are defined so that values greater than zero indicate that the fields decay exponentially along the +*x* and –*x* directions respectively, while values less than zero indicate that the fields grow exponentially. So the boundedness of the wave puts the following requirement on the materials.

$$\mu_1 > 0 \Rightarrow \mu_2 < 0 \quad \text{(TE Only)} \tag{3.49}$$

$$\varepsilon_1 > 0 \Rightarrow \varepsilon_2 < 0 \quad \text{(TM Only)} \tag{3.50}$$

Equations 3.45, 3.46 and 3.47 can be solved simultaneously for N, γ_1 and γ_2.

$$N_{eff}^2 = \varepsilon_1 \mu_2 \frac{(\mu_2/\mu_1) - (\varepsilon_2/\varepsilon_1)}{(\mu_2/\mu_1)^2 - 1} \tag{3.51}$$

$$\left(\frac{\gamma_1}{k_0}\right)^2 = \varepsilon_1 \mu_1 \frac{1 - (\varepsilon_2/\varepsilon_1)(\mu_2/\mu_1)}{(\mu_2/\mu_1)^2 - 1} \tag{3.52}$$

$$\left(\frac{\gamma_2}{k_0}\right)^2 = (\mu_2/\mu_1)^2 \left(\frac{\gamma_1}{k_0}\right)^2 \tag{3.53}$$

So for the TE solution to be support surface bound solutions, these three expressions must be greater than zero. Provided that the permeability in both regions is real, 3.53 is actually superfluous if 3.52 is satisfied. The lateral boundedness provided by 3.52 is an important distinction from 3.51 though; if the wave laterally propagates on both sides, the system is analogous to simple refraction. If it propagates along one direction and decays along another, the propagation along the surface would be analogous to the Goos-Hänchen effect (i.e., when a wave exhibits total internal reflection at an interface it is evanescent within the reflecting medium, but it can still propagate / change phase *along* the interface) [3]. These requirements can be satisfied two different ways.

$$(\mu_2/\mu_1) < -1 \tag{3.54}$$

$$(\mu_2/\mu_1) < (\varepsilon_2/\varepsilon_1) \tag{3.55}$$

$$(\mu_2/\mu_1)(\varepsilon_2/\varepsilon_1) < 1 \tag{3.56}$$

Or

$$-1 < (\mu_2/\mu_1) < 0 \tag{3.57}$$

$$(\mu_2/\mu_1) > (\varepsilon_2/\varepsilon_1) \tag{3.58}$$

$$(\mu_2/\mu_1)(\varepsilon_2/\varepsilon_1) > 1 \tag{3.59}$$

The TM conditions can be derived from duality by simply interchanging ε_2 and μ_2, and ε_1 and μ_1.

$$(\varepsilon_2 / \varepsilon_1) < -1 \tag{3.60}$$

$$(\varepsilon_2 / \varepsilon_1) < (\mu_2 / \mu_1) \tag{3.61}$$

$$(\varepsilon_2 / \varepsilon_1)(\mu_2 / \mu_1) < 1 \tag{3.62}$$

Or

$$-1 < (\varepsilon_2 / \varepsilon_1) < 0 \tag{3.63}$$

$$(\varepsilon_2 / \varepsilon_1) > (\mu_2 / \mu_1) \tag{3.64}$$

$$(\varepsilon_2 / \varepsilon_1)(\mu_2 / \mu_1) > 1 \tag{3.65}$$

The regions in (ε,μ)-space where surface waves are supported are shown in Figure 14. In this formulation, it is assumed that both the permittivity and permeability are real (the system is assumed to be lossless). The 3rd quadrant is representative of left-handed materials. The same sort of surface-bound waves are supported in left-handed systems as are supported in metallic systems, but in this case both polarizations are supported. This suggests that the Poynting vector circulation observed in left-handed materials [17],[22] is essentially identical to SP propagation in metallic-dielectric systems.

In Figure 14(a), it is shown that TM waves can be supported both in dielectric-metallic systems (2nd quadrant) and dielectic-left-handed systems (3rd quadrant). Likewise, TE waves can be supported in dielectric-left-handed systems or in systems with negative permeability materials (4th quadrant). Figure 14(b) shows how the region of propagation change when the second material is allowed to be slightly lossy. In the blue region, the TM mode is not supported even for a lossless system. In the region inside the teal-blue boundary, the TM mode is supported when the system is assumed to be entirely lossless. In the region inside the yellow-teal boundary, it is supported when $\mathrm{Im}(\varepsilon_2) \sim 0.1\mathrm{Re}(\varepsilon_2)$. In the red region, it is supported when $\mathrm{Im}(\varepsilon_2) \sim 0.2\mathrm{Re}(\varepsilon_2)$. These calculations are based on the following requirements.

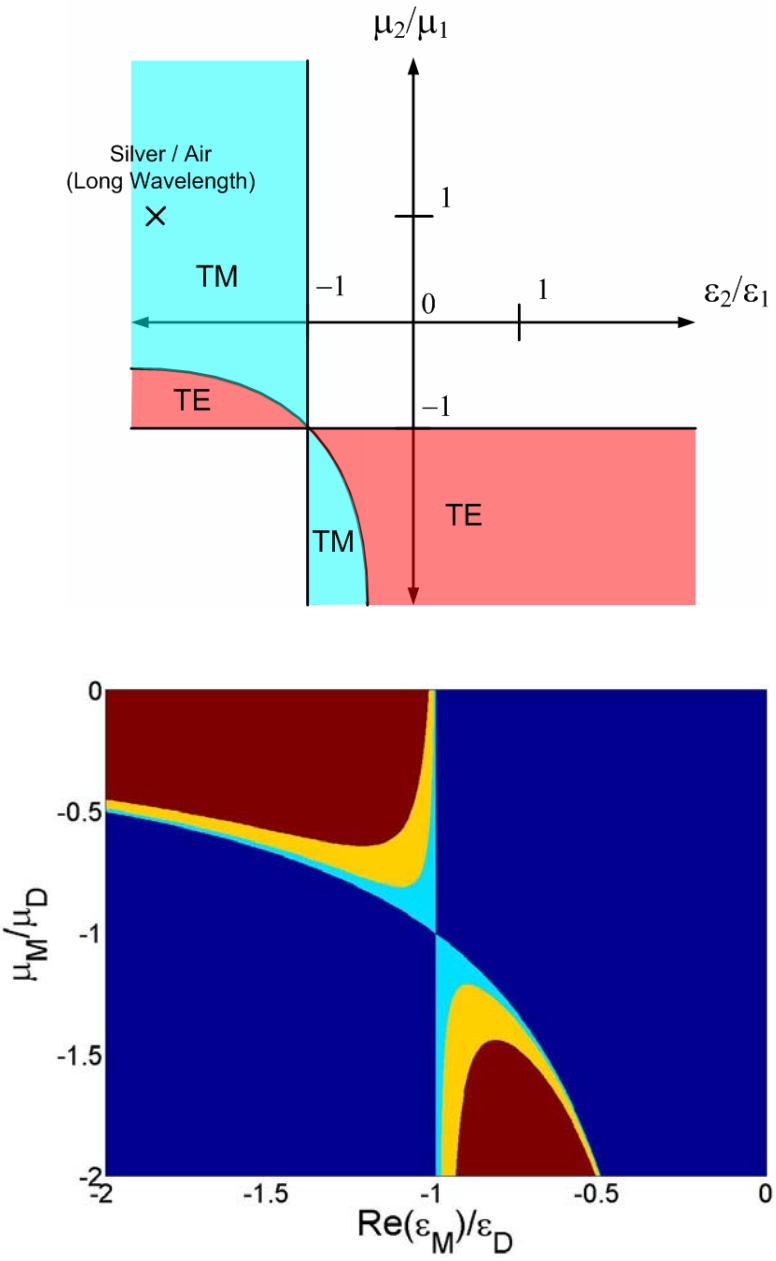

Figure 14: (a) Taken from [23], an illustration of the requirements for plasmon propagation at the interface between a dielectric material, medium 1 (ε_1, $\mu_1 > 0$) and some other material, medium 2 (ε_2, μ_2), for both TE and TM waves (b) focusing on the third quadrant and the TE wave, the dielectric constant of medium 2 is allowed to be complex, as described in the text.

$$\mathrm{Re}\left(N_{eff}^2\right) > 0 \tag{3.66}$$

$$\mathrm{Re}\left[\left(\frac{\gamma_1}{k_0}\right)^2\right] > 0 \tag{3.67}$$

$$\mathrm{Re}\left[\left(\frac{\gamma_2}{k_0}\right)^2\right] > 0 \tag{3.68}$$

The requirement that the squares are real implies that the complex values are within 45º of the real axis. Essentially this means the quantities are mostly real.

3.3.4 Poynting Vector Vorticity

In the previous sections the vorticity of the Poynting vector fields was discussed. It can easily be observed from Figure 11 that the vortices are asymmetric with a larger magnitude on the dielectric side than on the metal side. The asymmetry decreases as the wavelength decreases along with the magnitude of the metallic dielectric constant.

A further understanding of the nature of the Poynting vector vorticity can be gained by comparing the Poynting vector field to the electromagnetic fields and surface polarization charge distributions. These are shown for a single planar interface at four different free-space wavelengths in Figures 12 (λ_0 = 350 nm and 400 nm) and 13 (λ_0 = 500 nm and 650 nm). There appears to be a clear relationship between the Poynting vector and the polarization charge. The Poynting vector field is mostly tangential to the surface near the extrema of the polarization charge and mostly normal to the surface near the polarization charge nodes. Due to attenuation, the vortex will also be asymmetric in the direction of propagation; that is to say, it is stronger on end than the other. From a power flow point of view, this indicates that there is more power

circulating into the metal on one end than is circulating out at the other end, the difference clearly being lost to absorption in the metal region. Near resonance (the top part of Figure 12, λ_0 = 350 nm), the H_z field peaks appear slanted in the y-direction; this is indicative of a non-negligible imaginary component to γ_D. In other words, there is some phase shift in the direction normal to the interface.

While qualitative analysis is somewhat enlightening, understanding the mathematical basis for this vorticity is helpful. Vorticity is generally characterized by the curl of a vector field. The curl of the Poynting vector can be expressed as follows.

$$\nabla \times \vec{\mathrm{S}} = \nabla \times \left(\vec{E} \times \vec{H} \right) \tag{3.69}$$

$$\nabla \times \vec{\mathrm{S}} = \vec{E}\left(\nabla \cdot \vec{H} \right) - \vec{H}\left(\nabla \cdot \vec{E} \right) - \left(\vec{E} \cdot \nabla \right)\vec{H} + \left(\vec{H} \cdot \nabla \right)\vec{E} \tag{3.70}$$

Maxwell's equations can be used to simplify 3.70 significantly. First, the permeability is constant with respect to position, so $\mu_0 \nabla \cdot \vec{H} = \nabla \cdot \vec{B}$ is zero everywhere. Previously in expression 3.2, the polarization charge was related to the divergence of the electric field. The TM polarization state restricts the magnetic field to the z-direction – this will also help simplify the result.

$$\nabla \times \vec{\mathrm{S}} = -\hat{z} H_z \frac{\rho_P}{\varepsilon_0} - \hat{z}\left(E_x \partial_x H_z + E_y \partial_y H_z \right) + H_z \partial_z \vec{E} \tag{3.71}$$

All the fields are constant in the z-direction, so the last term disappears. Also, utilizing Ampere's law sets the middle terms to zero.

$$E_x \partial_x H_z + E_y \partial_y H_z = -i\omega\varepsilon\left(-E_x E_y + E_y E_x \right) = 0 \tag{3.72}$$

$$\nabla \times \vec{\mathbf{S}} = -\hat{z}\frac{\rho_P}{\varepsilon_0} H_z \tag{3.73}$$

Overall the curl of Poynting vector field is zero in both the metal and dielectric sides and is nonzero only on the interface as expressed above. The polarization charge density ρ_s and the magnetic field H_z will be roughly in phase (as discussed in Section 3.1), therefore $\nabla \times \mathbf{S}$ will tend to be either uniformly positive or negative. The sign of the vorticity dictates the sense of rotation (clockwise or counter-clockwise, which depends on the direction of propagation).

3.4 SINGLE ISOLATED SLIT

The analysis of a single-interface revealed quite a bit about surface plasmon propagation, but the real goal of our work is to study a metallic slit. Figures 12 and 13 illustrated the interaction between the charge, fields and Poynting vector for a single interface - a special, limiting case of the slit structure. Figure 15 takes the system at 400 nm free-space wavelength and introduces a second silver interface 100 nm and 25 nm away from the first one, illustrating what happens to the fields. The effect of coupling with a second interface has a similar effect to the reduction of the dielectric constant towards the resonant value ($\varepsilon_M = -\varepsilon_D$), the effective wavelength of propagation in the slit decreases and the propagation length decreases. In both cases, this effect arises from a larger portion of the power going into the metal region. In the uncoupled limit, the penetration depth for the fields on the air side is approximately 200 nm. This suggests that the coupling effects will become significant when the two interfaces are on the order of 100 nm

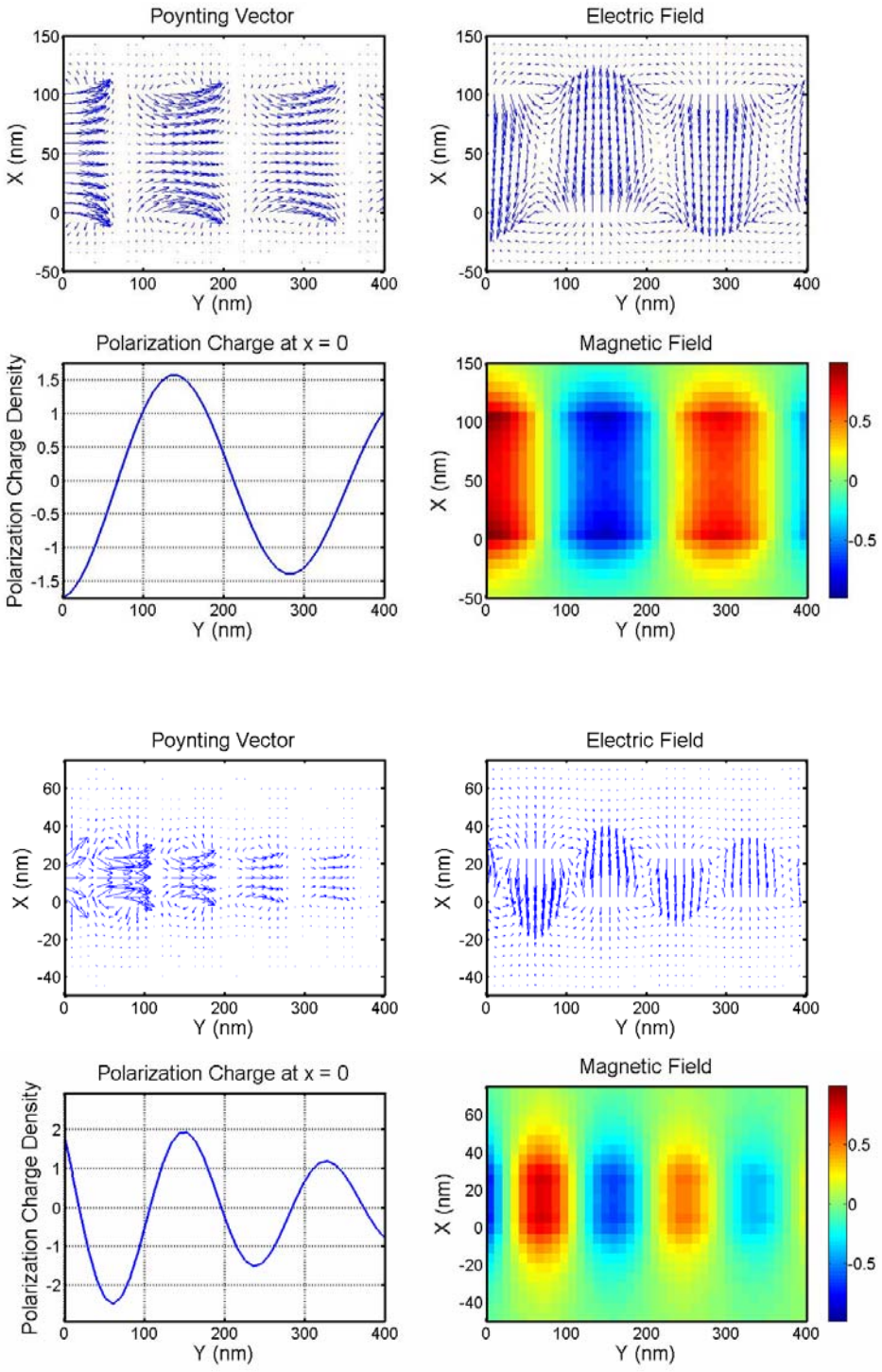

Figure 15: The instantaneous fields and polarization charge density at 400 nm free-space wavelength for a 100 nm slit (top) and a 25 nm slit (bottom).

apart; in other words, this will be when the interference between the two SP modes will become significant. The amplitude of the charge polarization also seems to increase as slit width decreases, a result predicted by both the coupling and effective index terms in expression 3.14. It was illustrated the polarization charge waves should be anti-symmetric for the 'symmetric' SP mode. It can be seen from the plots of the electric field that the anti-symmetric polarization charge profiles induce electric fields in the same direction, indicative of constructive interference between the two interfaces.

The previous discussion of power flow culminated in equations 3.43 and 3.44. From these results, the ratio of the power flowing in the metal to the power flowing in the slit region can be analytically calculated. A plot of this ratio over a range of wavelengths and slit widths is shown in Figure 16.

Figure 17 shows the angle of the time-averaged Poynting vector as a function of position around the slit (with the dimension of the x-axis normalized with respect to the slit width), calculated from 3.17-3.22. There is an inherent 180º rotation of the time-averaged Poynting vector; this overall shift is removed from the calculation by defining the angles with respect to the +y-direction in the slit region and with respect to the –y-direction in the metallic walls (as shown on the left-side of the figure). As discussed in Section 3.3.2, the angle on the metal side is constant with respect to position, while the angle on the air side changes substantially. For the sake of discussion, the tilt angle on the air side will refer to the angle at the interface (which is the location of maximum tilt). The top shows the result for 650 nm free-space wavelength and the bottom shows the result for 400 nm free-space wavelength. At the higher wavelength, the angle on the metal side ranges roughly from 8º to 6º as the slit width is decreased by a factor of 8 (from 200 nm to 25 nm). On the slit side this corresponds to max tilt angles of 0.46º and 0.35º

respectively. This is an interesting trend, because at smaller slit widths more power is flowing on the metal side (see Figure 16), so one would expect the amount of loss – characterized by the normal-component – to be proportionally larger. While the loss is increasing, the effective index and the propagation constant are also increasing substantially as slit width decreases. The expressions in 3.17-3.19 illustrate that the y-component of the time-averaged Poynting vector is proportional to the real part of the effective index, so the increase in the effective index results in an equivalent increase in the transverse component of the Poynting vector. The tilt angle is significantly larger for lower wavelengths, the wavelength dependence is also more complicated. At 400 nm wavelength, the angle varies from 13° at 200 nm slit width to to 12.9° at 100 nm slit width, then goes back up to about 14° at 25 nm slit width. The angle on the dielectric-side interface is approximately constant at 3.5° for all slit widths. Near resonance at low slit widths, the SP mode tends to be rather lossy - the imaginary part of the dielectric constant causes some sinusoidal variation in the power flow as a function of position in this regime. Overall, the variation with slit width is much less monotonic at 400 nm, as shown in parts (c) and (d) of Figure 17.

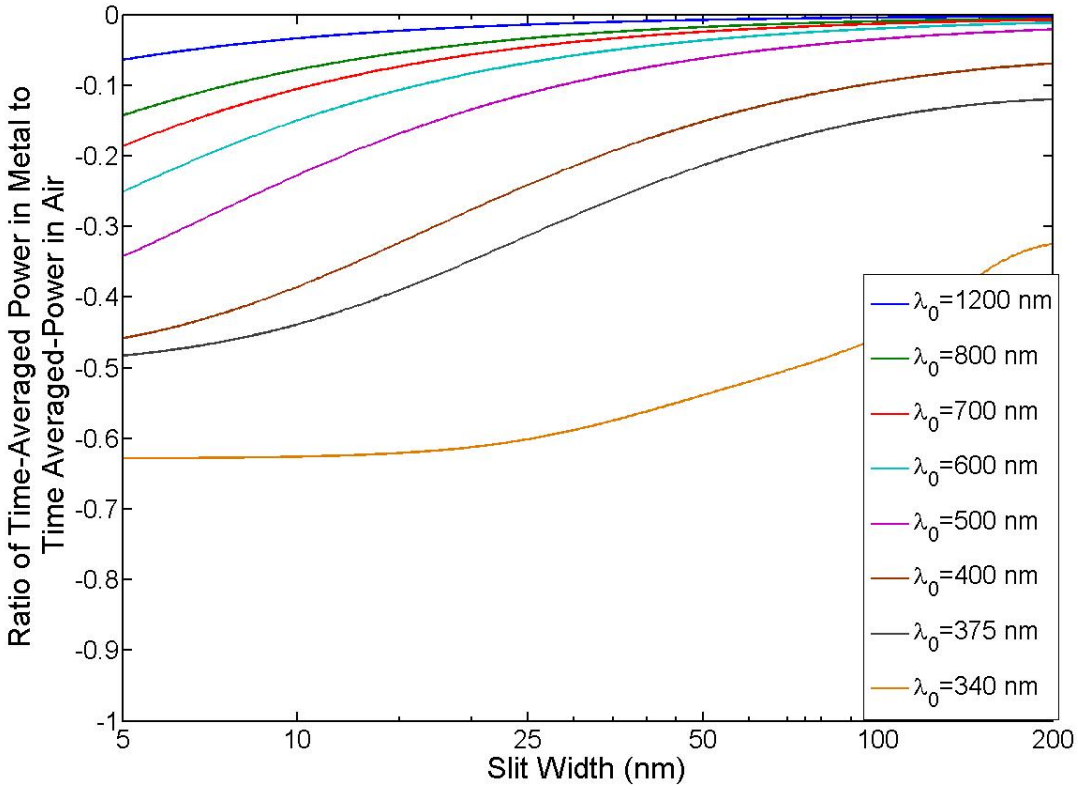

Figure 16: Ratio of the total power propagating in the metal (silver) region (backward propagating) to the total power propagating in the air filled slit region (forward propagating). A ratio greater than -1 indicates that the overall result is forward propagating. Taken from [23].

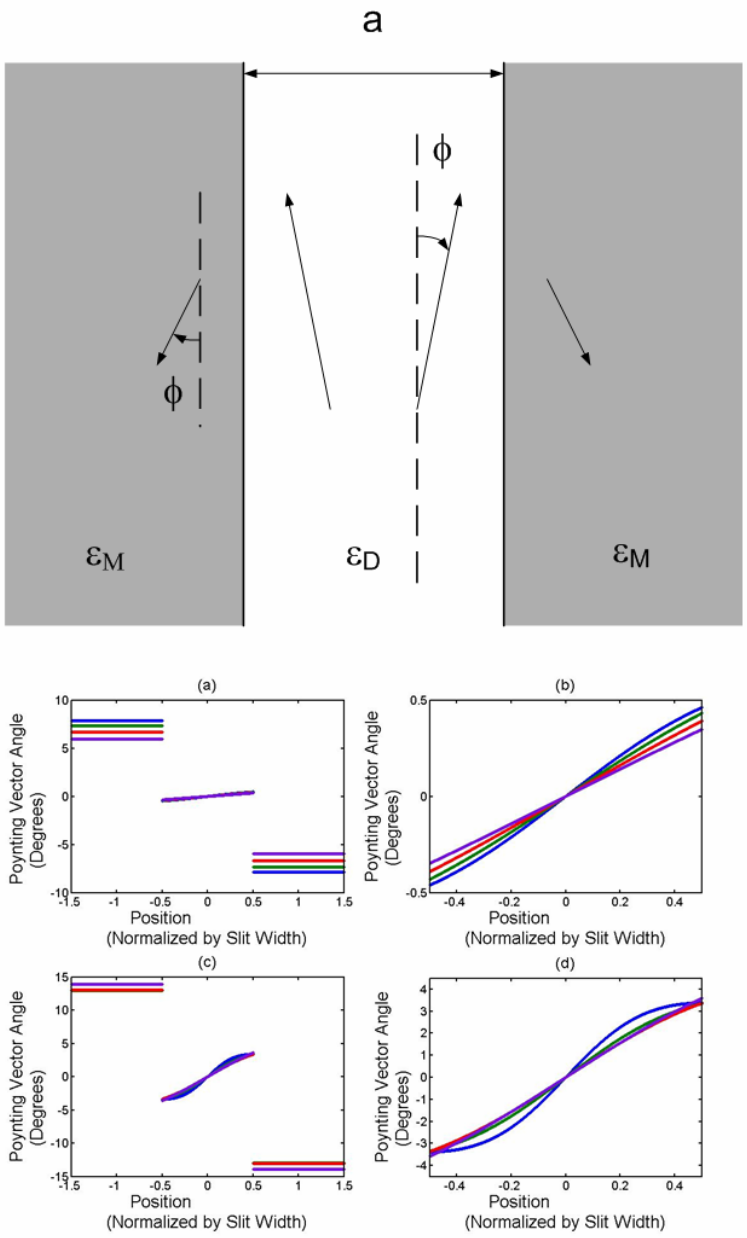

Figure 17: The orientation of the time-averaged Poynting vector in the metal (silver) and slit (air) regions, calculated at 650 nm wavelength [(a) and (b)] and 400 nm wavelength [(c) and (d)], and for slit width of 200 nm (blue), 100 nm (green), 50 nm (red), and 25 nm (violet). The right panels [(b) and (d)] show a magnified view of the slit region. Here the orientation is expressed as the tilt angle with respect to the +y direction in the slit region and to the –y direction in the metal region. (See top part), taken from [23].

4.0 APPLICATIONS

4.1 CHEMICAL SENSING

It has been analytically shown that the SP dispersion relationship and the resulting optical properties of a metallic slit structure are strongly dependent on the properties of the two materials that form the interface on which the SP propagates. Due to its localized nature the SP is most strongly affected by the material environment very near to the interface (within a hundred nanometers or so). By slightly altering that environment, relatively significant changes can be made to the quantitative optical properties of the system.

For example, consider a plasmonic system which exhibits a strong resonant reflection or transmission peak at a particular wavelength. That wavelength is determined by the strong wavelength dependence of the metallic dielectric constant in conjunction with the geometry. By coating the surface with a chemical that exhibits a dielectric constant different than the ambient dielectric, that resonant wavelength can be shifted drastically. A film that is one or two hundred nanometers thick – a relatively small amount of material - would almost completely change the dielectric constant felt by the SP on the dielectric side of the interface.

4.1.1 Resonant Conditions

The transmission profile of a metallic grating can be quite complex. Usually in periodic structures, the transmission peaks and dips can be related to certain resonances related to the periodicity. It has been argued in the previous chapters that in thick sub-wavelength gratings, light transmission is governed primarily by plasmonic modes. It seems reasonable then to ascribe the transmission peaks and dips to resonances associated with particular interfaces. This idea has been explored in other works [25].

Consider a single, rectangular metallic island within a larger grating structure. Its dimensions will be given by the thickness of the grating (t) and the total grating period minus the slit width (see Figure 18; L + a = P). One side is facing a semi-infinite air region, one side is facing a semi-infinite quartz substrate (practically semi-infinite, more than 100 μm and certainly much greater than λ_0), and two sides that look similar to the slit waveguides studied in previous chapters. Assume that the front and back faces can be modeled as isolated interfaces (with dielectric on one side and metal on another) and that the sides can be approximated as the infinitely long slit structures described in Chapters 2 and 3. It is possible that the effect of corners could cause distortion in these modes, but because the penetration of the plasmon into the metal is only on the order of 20 nm or so, the overlap between the two modes at the corners is small. This is – of course – a highly idealized structure, but as fabrication techniques become more sophisticated, well-defined geometries such as this are becoming easier to fabricate on slit width scales of less than 100 nm.

Sharp corners on the structure tend to make the analysis more difficult. From the perspective of a wave propagating on the surface, there are a number of sharp discontinuities

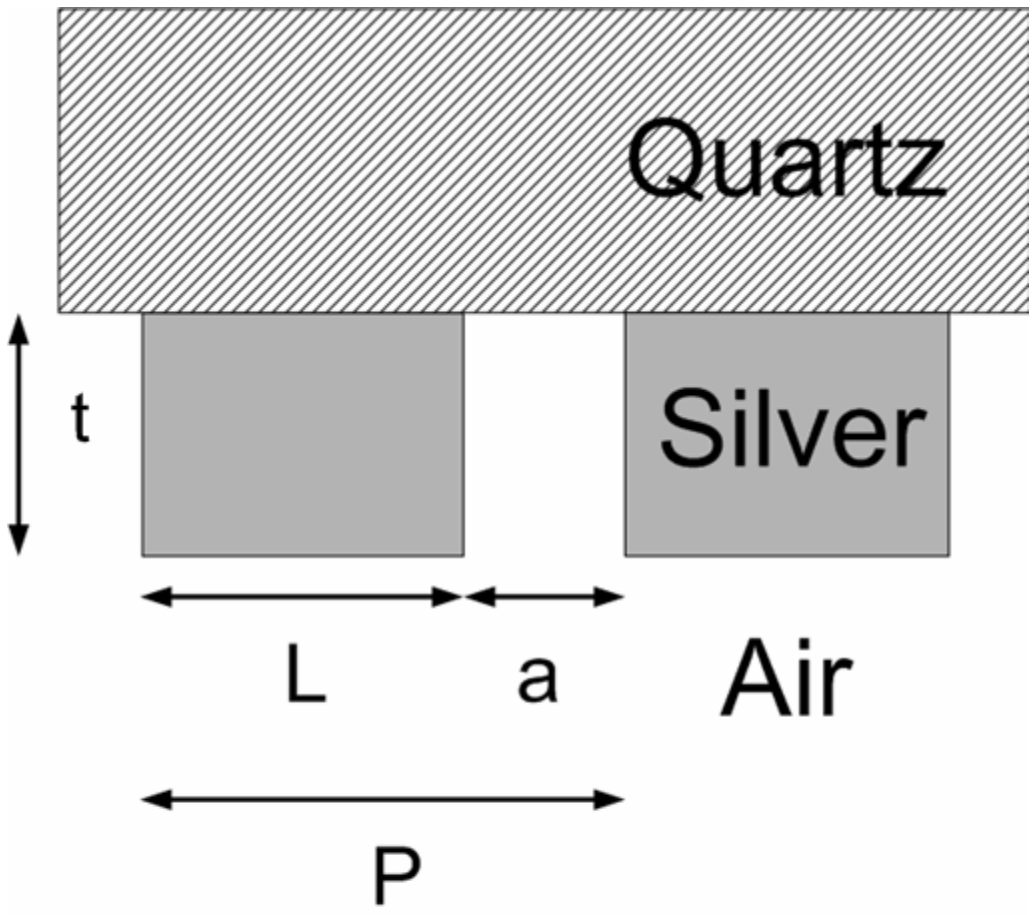

Figure 18: A sketch of two periods of a silver slit array on a quartz substrate. The structure has a periodicity of P, slit width a, and silver thickness t.

which can form the fixed points of standing waves. For example, one could envision a standing wave associated with just the front or back face of the metal, with the charge bound to zero at the corners. A resonance condition can be derived for such a mode.

$$\frac{m}{2}\lambda_0 = N_{Front}L \tag{4.1}$$

If the front face is approximately equal to an isolated, single interface, then N_{Front} is known.

$$N_{Front} \approx \sqrt{\frac{\varepsilon_M \varepsilon_D}{\varepsilon_M + \varepsilon_D}} \tag{4.2}$$

Likewise, a resonant mode circulating all the way around the island could be considered. A round-trip would require a 2πm phase shift (some multiple of a full wavelength) to satisfy the phase matching condition.

$$m\lambda_0 = (N_{Front} + N_{Back})L + 2N_{Slit}t \tag{4.3}$$

If the dimensions of the structure are on the order of a single wavelength, it is very likely that the lower order modes – m = 1 or 2 – will be the most strongly excited. It has been proposed that a (m = 2) resonant mode occurs at the high transmission peak in this sort of structure [25]**.** However these resonant conditions are not as simple as they appear, due to the dependence of the effective index with wavelength. Even the simple single interface effective index is dependent on the dielectric function of the metal, which tends to be a strong function of frequency.

Suppose that a transmission peak is caused by a particular resonance and one wanted to consider the wavelength shift caused by a small perturbation of some general parameter, x,

which may cause a corresponding change in the effective indices. For example, slightly changing the background dielectric from $\varepsilon_D = 1$ to $\varepsilon_D = 1.1$ would cause a corresponding change in N_{Front} and N_{Slit}.

$$m\lambda_0 = (N_{Front}(\lambda_0, x) + N_{Back}(\lambda_0, x))L + 2N_{Slit}(\lambda_0, x)t \tag{4.4}$$

$$m\delta\lambda_0 = \left[\left(\frac{\partial N_{Front}}{\partial \lambda_0} + \frac{\partial N_{Back}}{\partial \lambda_0}\right)L + 2\frac{\partial N_{Slit}}{\partial \lambda_0}t\right]\delta\lambda_0 + \left[\left(\frac{\partial N_{Front}}{\partial x} + \frac{\partial N_{Back}}{\partial x}\right)L + 2\frac{\partial N_{Slit}}{\partial x}t\right]\delta x \tag{4.5}$$

$$\delta\lambda_0 = \frac{\left[\left(\frac{\partial N_{Front}}{\partial x} + \frac{\partial N_{Back}}{\partial x}\right)L + 2\frac{\partial N_{Slit}}{\partial x}t\right]}{m - \left[\left(\frac{\partial N_{Front}}{\partial \lambda_0} + \frac{\partial N_{Back}}{\partial \lambda_0}\right)L + 2\frac{\partial N_{Slit}}{\partial \lambda_0}t\right]}\delta x \tag{4.6}$$

Alternatively, this expression can be written in terms of the propagation constant, β.

$$2\pi m = (\beta_{Front}(\lambda_0, x) + \beta_{Back}(\lambda_0, x))L + 2\beta_{Slit}(\lambda_0, x)t \tag{4.7}$$

$$0 = \left[\left(\frac{\partial \beta_{Front}}{\partial \lambda_0} + \frac{\partial \beta_{Back}}{\partial \lambda_0}\right)L + 2\frac{\partial \beta_{Slit}}{\partial \lambda_0}t\right]\delta\lambda_0 + \left[\left(\frac{\partial \beta_{Front}}{\partial x} + \frac{\partial \beta_{Back}}{\partial x}\right)L + 2\frac{\partial \beta_{Slit}}{\partial x}t\right]\delta x \tag{4.8}$$

$$\delta\lambda_0 = -\frac{\left[\left(\frac{\partial \beta_{Front}}{\partial x} + \frac{\partial \beta_{Back}}{\partial x}\right)L + 2\frac{\partial \beta_{Slit}}{\partial x}t\right]}{\left[\left(\frac{\partial \beta_{Front}}{\partial \lambda_0} + \frac{\partial \beta_{Back}}{\partial \lambda_0}\right)L + 2\frac{\partial \beta_{Slit}}{\partial \lambda_0}t\right]}\delta x \tag{4.9}$$

In 4.9, β refers only to the real part of the propagation constant (which determines the phase shift per unit length). Unfortunately, expressions for the effective index tend to be very complicated. In the slit case, the expression is implicit, making it difficult to analytically calculate a derivative of any parameter. This expression may be useful if the magnitude or even signs of the expressions are known from physical intuition. Except for regions of anomalous dispersion,

effective index tends to decrease with respect to wavelength. This expression will be employed towards the end of the following section.

4.1.2 Sensitivity Calculations and Approximations

Consider the slit structure discussed in the first few chapters, but with a thin (a few nanometers thick) analyte layer added to the front side and inside the slit region. The dielectric constant of this material will be different than air, so it will change the effective index of three out of four sides, change the plasmon resonances and change the overall transmission spectrum. In this section, a slit with a thin layer of dielectric deposited on the metal-air interface will be modeled analytically. Afterwards more useful, approximate expressions will be derived.

A semi-infinite slit in silver of width "a" is filled with a material of dielectric constant ε_D, and is coated on the sides by a film of thickness d and dielectric constant ε_F. There is no restriction on the thickness other than it must be less than half the slit width. This geometry is illustrated in Figure 19.

Based on symmetry the following form is assumed for the z-component of the magnetic field in the slit region (for reasons discussed above, we focus on the symmetric SP mode).

$$H_z^{(D)}(x,y) = H_0^{(D)}(e^{-\gamma_D(x-d)} + e^{\gamma_D(x-a+d)})e^{i\beta y},\ d \le x \le a-d \qquad (4.10)$$

$H_0^{(D)}$ can be defined arbitrarily and $\gamma_D = \sqrt{\beta^2 - \varepsilon_D k_0^2}$. The solution in the dielectric layer likewise can be expressed as a sum of decaying and growing exponentials.

$$H_z^{(F)}(x,y) = (Ae^{\gamma_F x} + Be^{-\gamma_F(x-d)})e^{i\beta y},\ 0 \le x \le d \qquad (4.11)$$

$$\gamma_F = \sqrt{\beta^2 - \varepsilon_F k_0^2} \tag{4.12}$$

Due to phase matching at the interfaces, β will be the same in each layer. The form of the exponentials are set in anticipation of solving for the boundary conditions at x = 0 and x = a – d and based on the assumed form of the solution. Finally, the fields should die off as x goes to negative infinity.

$$H_z^{(M)}(x,y) = H_0^{(M)} e^{\gamma_M x} e^{i\beta y}, \quad x \le 0 \tag{4.13}$$

$$\gamma_M = \sqrt{\beta^2 - \varepsilon_M k_0^2} \tag{4.14}$$

$$H_z^{(M)}(0,y) = H_z^{(F)}(0,y) \tag{4.15}$$

$$H_z^{(F)}(d,y) = H_z^{(D)}(d,y) \tag{4.16}$$

$$H_0^{(M)} = A + Be^{\gamma_F d} \tag{4.17}$$

$$H_0^{(D)}(1 + e^{\gamma_D(2d-a)}) = Ae^{\gamma_D d} + B \tag{4.18}$$

The tangential component of the electric field (E_y) must also be continuous. E_y can be expressed in terms of H_z.

$$E_y(x,y) = \frac{1}{i\omega\varepsilon}\frac{\partial H_z}{\partial x}(x,y) \tag{4.19}$$

Thus the above expressions for H_z can be converted into corresponding expressions for E_y in each region

$$E_y^{(D)}(x,y) = -\frac{\gamma_D}{i\omega\varepsilon_D} H_0^{(D)}(e^{-\gamma_D(x-d)} - e^{\gamma_D(x-a+d)})e^{i\beta y} \tag{4.20}$$

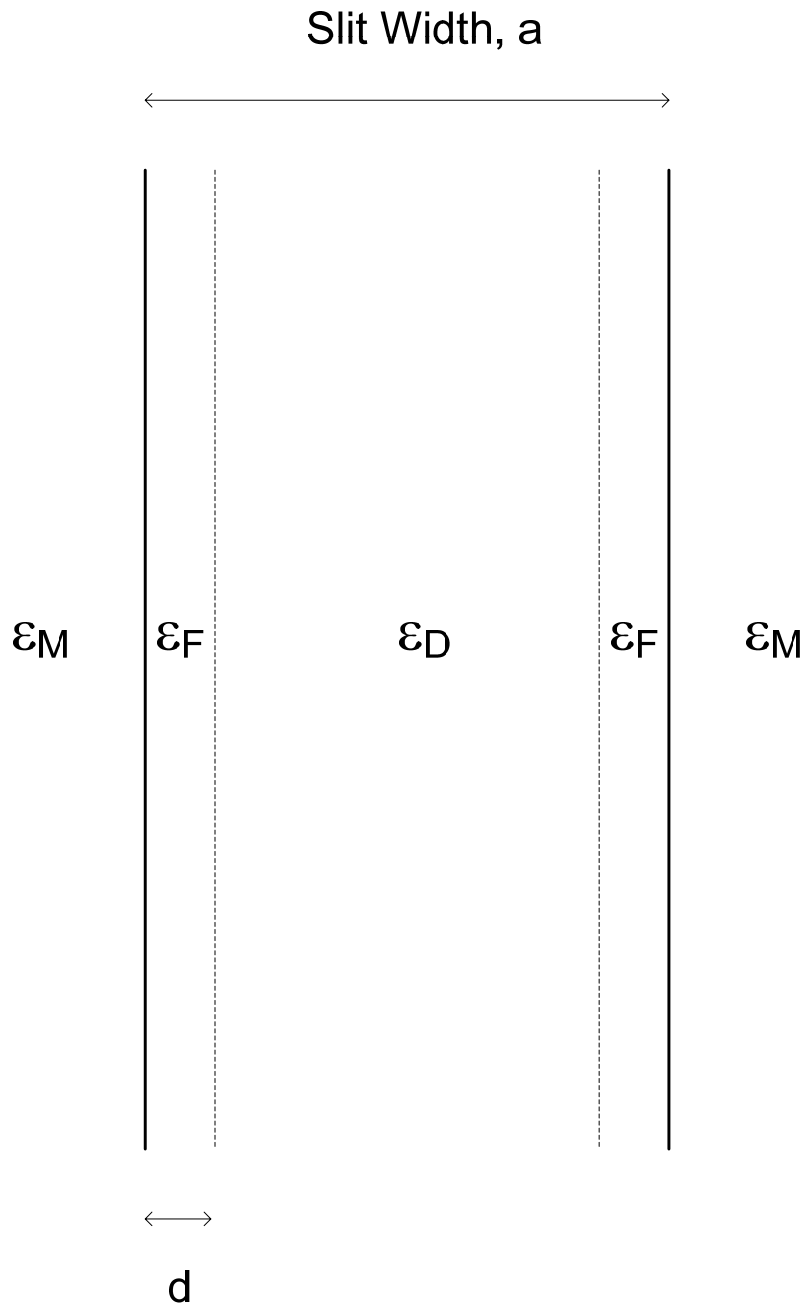

Figure 19: Sketch of 'modified' single infinite slit geometry. The metallic slit walls (designated by ε_M) are a distance a apart and covered with a thin film (ε_F) of some material. The rest of the slit is filled with air or water (ε_D).

$$E_y^{(F)}(x,y) = \frac{\gamma_F}{i\omega\varepsilon_F}(Ae^{\gamma_F x} - Be^{-\gamma_F (x-d)})e^{i\beta y},\ 0 \le x \le d \tag{4.21}$$

$$E_y^{(M)}(x,y) = \frac{\gamma_M}{i\omega\varepsilon_M} H_0^{(M)} e^{\gamma_M x} e^{i\beta y},\quad x \le 0 \tag{4.22}$$

Continuity must be satisfied at the interfaces x = 0 and x = d.

$$E_y^{(M)}(0,y) = E_y^{(F)}(0,y) \tag{4.23}$$

$$E_y^{(F)}(d,y) = E_y^{(D)}(d,y) \tag{4.24}$$

The first set of boundary conditions (requiring continuity of H_z) will be as follows,

$$H_0^{(M)} \frac{\gamma_M}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_M} = A - Be^{\gamma_F d} \tag{4.25}$$

$$-H_0^{(D)}(1 - e^{\gamma_D(2d-a)})\frac{\gamma_D}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_D} = Ae^{\gamma_F d} - B \tag{4.26}$$

The goal is now to solve for the propagation constant, β. Combining 4.17 and 4.25 eliminates $H_0^{(M)}$

$$\frac{\gamma_M}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_M}(A + Be^{\gamma_F d}) = A - Be^{\gamma_F d} \tag{4.27}$$

$$\left(\frac{\gamma_M}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_M} - 1\right)A + \left(\frac{\gamma_M}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_M} + 1\right)e^{\gamma_F d}B = 0 \tag{4.28}$$

A relationship between A and B comes out of 4.28. For notational simplicity, define α such that $B = \alpha A$.

$$\alpha = \frac{\left(1 - \frac{\gamma_M}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_M}\right)}{\left(1 + \frac{\gamma_M}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_M}\right)} e^{-\gamma_F d} \tag{4.29}$$

Likewise, $H_0^{(D)}$ can be eliminated with 4.18 and 4.26.

$$\frac{(1 - e^{\gamma_D(2d-a)})}{(1 + e^{\gamma_D(2d-a)})}\frac{\gamma_D}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_D} = \frac{\alpha - e^{\gamma_F d}}{\alpha + e^{\gamma_F d}} \tag{4.30}$$

This result can be put in a more convenient form by substituting in the definition of α given above, multiplying the top and bottom of the right-hand side by $(1 + \frac{\gamma_M}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_M})$, and multiplying both sides by γ_F/ε_F.

$$\tanh\left[\gamma_D\left(\frac{a}{2} - d\right)\right]\frac{\gamma_D}{\varepsilon_D} = \frac{\left(\frac{\gamma_F}{\varepsilon_F} - \frac{\gamma_M}{\varepsilon_M}\right)e^{-\gamma_2 d} - \left(\frac{\gamma_F}{\varepsilon_F} + \frac{\gamma_M}{\varepsilon_M}\right)e^{\gamma_2 d}}{\left(\frac{\gamma_F}{\varepsilon_F} - \frac{\gamma_M}{\varepsilon_M}\right)e^{-\gamma_2 d} + \left(\frac{\gamma_F}{\varepsilon_F} + \frac{\gamma_M}{\varepsilon_M}\right)e^{\gamma_2 d}} \tag{4.31}$$

It is somewhat difficult to see directly, but it turns out that the only unknown in this expression is the propagation constant β. Everything is either a function of β or a constant. Note that this expression should reduce to the original SIS expression as d goes to zero. This can be shown with a few steps of algebra.

$$\tanh\left[\gamma_D \frac{a}{2}\right]\frac{\gamma_D}{\varepsilon_D} = \frac{\left(\frac{\gamma_F}{\varepsilon_F} - \frac{\gamma_M}{\varepsilon_M}\right) - \left(\frac{\gamma_F}{\varepsilon_F} + \frac{\gamma_M}{\varepsilon_M}\right)}{\left(\frac{\gamma_F}{\varepsilon_F} - \frac{\gamma_M}{\varepsilon_M}\right) + \left(\frac{\gamma_F}{\varepsilon_F} + \frac{\gamma_M}{\varepsilon_M}\right)} \tag{4.32}$$

$$\tanh\left[\gamma_D \frac{a}{2}\right]\frac{\gamma_D}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_D} = -\frac{\gamma_M}{\gamma_F}\frac{\varepsilon_F}{\varepsilon_M} \tag{4.33}$$

$$\frac{\gamma_D}{\varepsilon_D}\tanh\left(\gamma_D \frac{a}{2}\right) = -\frac{\gamma_M}{\varepsilon_M} \tag{4.34}$$

The result in 4.34 matches the eigenvalue equation for a SIS with no dielectric layer on the sidewalls (2.28).

If the front and back sides are assumed to be isolated metal-dielectric interfaces and the aforementioned model for the SIS is used in the slit region, the resonance condition can – in principle – be solved exactly for the resonance wavelength.

$$\lambda_0 = \frac{L}{2}\left[N_{Front}(\lambda_0, d) + N_{Back}(\lambda_0, d)\right] + N_{Slit}(\lambda_0, a, d)t \tag{4.35}$$

The strategy for calculating the peak shift based on this equation is as follows.

(1) Make an initial guess for the resonant wavelength of the uncoated structure; half the perimeter of the metallic island (L + t) is a good starting point

(2) At the wavelength of interest, calculate the effective indices based on the geometry (setting d = 0 for the uncoated structure)

(3) Use these values to calculate the right-hand side of 4.62, which is equal to the new guess for the resonant wavelength

(4) Iterate through steps (1) through (3) until the solution converges (say the updated guess is only different from the previous guess by 1% or less) or if it appears that the solution is not converging (if it exceeds a certain large value or if it has gone through more than 100 iterations without convergence)

(5) Repeat steps (1) through (4) but for the coated structure (let d equal the thickness of the analyte); use the result for the uncoated structure as an initial guess

This approach has at least one obvious flaw; not only does the effective index have to be calculated numerically for the coated structure, three of these calculations must be done for each iteration. If we assume that the result will converge within ten iterations, that means that the effective index eigenvalue equation will have to be solved 30 times for a single data point (corresponding to particular values of period, slit width, silver thickness, and analyte thickness). This might be reasonable if the peak shift for a specific geometry is needed, but from a design perspective it could be very troubling.

The eigenvalue equation in 4.31 is exact in the sense that it is derived directly from Maxwell's Equations (with a few extra assumptions, such as the quadrupolar nature of the resonance and ignoring the coupling at the corners). However, it is also implicit and must be solved numerically. An explicit relationship for the sensitivity and peak shift– even an approximate one – would be easier to evaluate and may provide greater intuition about the problem.

An approximate relationship for the peak shift can be calculated using the expression derived in 4.6. But in order to use this equation, a perturbation parameter (generically referred to

as x in the expression) has to be selected. We are interested in examining the peak shift caused by adding a small chemical layer with a different index of refraction than the background dielectric; so the obvious choice would be to pick the analyte thickness as the perturbation parameter. The problem with this choice is that thickness does not appear anywhere in the unperturbed eigenvalue equations; so this route would require making a first order approximation from the exact expression in 4.31, which is rather unwieldy. Physically, chemically modifying the surface on the air side would slightly change the effective dielectric constant ε_D seen by the plasmon wave. The slit side dielectric constant *does* appear in the unperturbed expression.

$$\delta\lambda_0 = \frac{\left[\left(\frac{\partial N_{Front}}{\partial \varepsilon_D} + \frac{\partial N_{Back}}{\partial \varepsilon_D}\right)L + 2\frac{\partial N_{Slit}}{\partial \varepsilon_D}t\right]}{m - \left[\left(\frac{\partial N_{Front}}{\partial \lambda_0} + \frac{\partial N_{Back}}{\partial \lambda_0}\right)L + 2\frac{\partial N_{Slit}}{\partial \lambda_0}t\right]}\delta\varepsilon_D \tag{4.36}$$

This expression should be modified a bit. First of all, there is strong evidence that peak transmission corresponds to a quadrupolar resonance, m = 2. For this choice of parameter, the expression in 4.36 is not entirely correct. A different effective dielectric constant is observed in the front and slit regions, so there is no general $\delta\varepsilon_D$ which applies to all of the sides. Also, the chemical layer is not deposited on the side of the metal island that is sitting on the substrate – the 'back' – so the effective dielectric constant on this side will not change.

$$\delta\lambda_0 = \frac{\frac{1}{2}\frac{\partial N_{Front}}{\partial \varepsilon_D^{Front}}L\delta\varepsilon_D^{Front} + \frac{\partial N_{Slit}}{\partial \varepsilon_D^{Slit}}t\delta\varepsilon_D^{Slit}}{1 - \frac{\partial N_{Slit}}{\partial \lambda_0}t - \left(\frac{\partial N_{Front}}{\partial \lambda_0} + \frac{\partial N_{Back}}{\partial \lambda_0}\right)\frac{L}{2}} \tag{4.37}$$

The version using the propagation constant instead of the effective index is also useful.

$$\delta\lambda_0 = -\frac{\frac{1}{2}\frac{\partial\beta_{Front}}{\partial\varepsilon_D^{Front}} L\delta\varepsilon_D^{Front} + \frac{\partial\beta_{Slit}}{\partial\varepsilon_D^{Slit}} t\delta\varepsilon_D^{Slit}}{\frac{\partial\beta_{Slit}}{\partial\lambda_0} t + \left(\frac{\partial\beta_{Front}}{\partial\lambda_0} + \frac{\partial\beta_{Back}}{\partial\lambda_0}\right)\frac{L}{2}} \tag{4.38}$$

The eigenvalue equation given in 2.28 is implicit and not trivial to work with, but it is possible to find exact expressions for the partial derivatives required in 4.37. For simplicity, use normalized versions of 2.17, 2.18, and 2.28, which are determined simply by dividing the decay constants by k_0.

$$\tilde{\gamma}_D \equiv \frac{\gamma_D}{k_0} \qquad \tilde{\gamma}_M \equiv \frac{\gamma_M}{k_0} \tag{4.39}$$

$$N^2 = \tilde{\gamma}_D^2 + \varepsilon_D = \tilde{\gamma}_M^2 + \varepsilon_M \tag{4.40}$$

$$\tilde{\gamma}_D \tanh\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right] = -\frac{\varepsilon_D}{\varepsilon_M}\tilde{\gamma}_M \tag{4.41}$$

The next step is to take the derivative of both sides of 4.40 and 4.41. First, find the derivative with respect to the dielectric constant.

$$N\frac{\partial N}{\partial\varepsilon_D} = \tilde{\gamma}_D \frac{\partial\tilde{\gamma}_D}{\partial\varepsilon_D} + \frac{1}{2} = \tilde{\gamma}_M \frac{\partial\tilde{\gamma}_M}{\partial\varepsilon_D} \tag{4.42}$$

In taking the derivative of 4.41, the following properties of the hyperbolic tangent are useful.

$$\frac{d}{dx}\tanh[x] = \frac{1}{\cosh^2[x]} = 1 - \tanh^2[x] \tag{4.43}$$

The rest is simply a matter of taking derivatives and substituting from 4.35 to get rid of expressions involving the derivatives of γ_D and γ_M. The strategy will be to solve in terms of $\frac{\partial \tilde{\gamma}_D}{\partial \varepsilon_D}$, then substitute in for $\frac{\partial N}{\partial \varepsilon_D}$.

$$\frac{\partial \tilde{\gamma}_D}{\partial \varepsilon_D}\left\{\tanh\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]+\tilde{\gamma}_D \frac{k_0 a}{2}\left(1-\tanh^2\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]\right)\right\} = -\frac{1}{\varepsilon_M}\tilde{\gamma}_M - \frac{\varepsilon_D}{\varepsilon_M}\frac{\partial \tilde{\gamma}_M}{\partial \varepsilon_D} \tag{4.44}$$

$$\frac{\partial \tilde{\gamma}_D}{\partial \varepsilon_D}\left\{\tanh\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]+\tilde{\gamma}_D \frac{k_0 a}{2}\left(1-\tanh^2\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]\right)+\frac{\varepsilon_D}{\varepsilon_M}\frac{\tilde{\gamma}_D}{\tilde{\gamma}_M}\right\} = -\frac{1}{\varepsilon_M}\tilde{\gamma}_M - \frac{\varepsilon_D}{\varepsilon_M}\frac{1}{2\tilde{\gamma}_M} \tag{4.45}$$

$$\frac{\partial \tilde{\gamma}_D}{\partial \varepsilon_D} = -\frac{\frac{\varepsilon_D}{\varepsilon_M}\frac{1}{2\tilde{\gamma}_M}+\frac{1}{\varepsilon_M}\tilde{\gamma}_M}{\tanh\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]+\tilde{\gamma}_D \frac{k_0 a}{2}\left(1-\tanh^2\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]\right)-\left(\frac{\varepsilon_D}{\varepsilon_M}\right)^2 \tanh^{-1}\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]} \tag{4.46}$$

$$\frac{\partial N}{\partial \varepsilon_D} = -\frac{\tilde{\gamma}_D}{N}\frac{\frac{\varepsilon_D}{\varepsilon_M}\frac{1}{2\tilde{\gamma}_M}+\frac{1}{\varepsilon_M}\tilde{\gamma}_M}{\tanh\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]+\tilde{\gamma}_D \frac{k_0 a}{2}\left(1-\tanh^2\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]\right)-\left(\frac{\varepsilon_D}{\varepsilon_M}\right)^2 \tanh^{-1}\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]}+\frac{1}{2N} \tag{4.47}$$

The approach for finding the derivative with respect to λ_0 is similar.

$$N\frac{\partial N}{\partial \lambda_0} = \tilde{\gamma}_D \frac{\partial \tilde{\gamma}_D}{\partial \lambda_0} = \tilde{\gamma}_M \frac{\partial \tilde{\gamma}_M}{\partial \lambda_0} + \frac{1}{2}\frac{\partial \varepsilon_M}{\partial \lambda_0} \tag{4.48}$$

$$\frac{\partial \tilde{\gamma}_D}{\partial \lambda_0}\tanh\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]+\tilde{\gamma}_D \frac{\partial}{\partial \lambda_0}\tanh\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right] = -\varepsilon_D\left[\frac{1}{\varepsilon_M}\frac{\partial \tilde{\gamma}_M}{\partial \lambda_0}-\frac{\tilde{\gamma}_M}{\varepsilon_M^2}\frac{\partial \varepsilon_M}{\partial \lambda_0}\right] \tag{4.49}$$

$$\frac{\partial \tilde{\gamma}_D}{\partial \lambda_0}\tanh\left[\tilde{\gamma}_D \frac{k_0 a}{2}\right]+a\pi\left(1-\tanh^2[\pi\tilde{\gamma}_D \frac{a}{\lambda_0}]\right)\left(\frac{\tilde{\gamma}_D}{\lambda_0}\frac{\partial \tilde{\gamma}_D}{\partial \lambda_0}-\frac{\tilde{\gamma}_D^2}{\lambda_0^2}\right) = -\varepsilon_D\left[\frac{1}{\varepsilon_M}\frac{\partial \tilde{\gamma}_M}{\partial \lambda_0}-\frac{\tilde{\gamma}_M}{\varepsilon_M^2}\frac{\partial \varepsilon_M}{\partial \lambda_0}\right] \tag{4.50}$$

In this case, it is more convenient to substitute $\frac{\partial N}{\partial \lambda_0}$ in directly.

$$\frac{\partial N}{\partial \lambda_0}\left\{\frac{N}{\widetilde{\gamma}_D}\tanh\left[\widetilde{\gamma}_D\,\frac{k_0 a}{2}\right]+\frac{N}{\lambda_0}a\pi\left(1-\tanh^2[\pi\widetilde{\gamma}_D\,\frac{a}{\lambda_0}]\right)+\frac{\varepsilon_D}{\widetilde{\gamma}_M\,\varepsilon_M}N\right\}= \\ \frac{\varepsilon_D}{\varepsilon_M}\frac{\partial \varepsilon_M}{\partial \lambda_0}\left(\frac{1}{2\widetilde{\gamma}_M}+\frac{\widetilde{\gamma}_M}{\varepsilon_M}\right)+\frac{\widetilde{\gamma}_D^2}{\lambda_0^2}a\pi\left(1-\tanh^2[\pi\widetilde{\gamma}_D\,\frac{a}{\lambda_0}]\right) \tag{4.51}$$

$$\frac{\partial N}{\partial \lambda_0}=\frac{1}{N}\frac{\frac{\varepsilon_D}{\varepsilon_M}\frac{\partial \varepsilon_M}{\partial \lambda_0}\left(\frac{1}{2\widetilde{\gamma}_M}+\frac{\widetilde{\gamma}_M}{\varepsilon_M}\right)+\frac{\widetilde{\gamma}_D^2}{\lambda_0^2}a\pi\left(1-\tanh^2\left[\widetilde{\gamma}_D\,\frac{k_0 a}{2}\right]\right)}{\frac{1}{\widetilde{\gamma}_D}\tanh\left[\widetilde{\gamma}_D\,\frac{k_0 a}{2}\right]+\frac{1}{\lambda_0}a\pi\left(1-\tanh^2\left[\widetilde{\gamma}_D\,\frac{k_0 a}{2}\right]\right)+\frac{\varepsilon_D}{\widetilde{\gamma}_M\,\varepsilon_M}} \tag{4.52}$$

Everything in 4.52 can be calculated explicitly from the design parameters of the problem; if the metallic dielectric function is known over a range of wavelengths [6], its slope can be approximated without too much trouble.

Approximating the change in the dielectric constant is more difficult, simply because the idea of an "effective dielectric constant" is a little physically ambiguous. One way to define it is in terms of energy density. The energy density of the electromagnetic field can be defined in a piece-wise fashion over the slit region itself and in the region of the analyte coating.

$$U(x,y)=\begin{cases}\frac{1}{2}\mu_0\left|\vec{H}(x,y)\right|^2+\frac{1}{2}\varepsilon_D\left|\vec{E}(x,y)\right|^2, & d>x>a-d \\ \frac{1}{2}\mu_0\left|\vec{H}(x,y)\right|^2+\frac{1}{2}\varepsilon_F\left|\vec{E}(x,y)\right|^2, & 0\geq x\geq d,\quad a-d\geq x\geq a\end{cases} \tag{4.53}$$

Alternatively, the idea of a single effective dielectric constant for the entire slit region can be utilized.

$$U_{Eff}(x,y)=\frac{1}{2}\mu_0\left|\vec{H}(x,y)\right|^2+\frac{1}{2}\varepsilon_{Eff}\left|\vec{E}(x,y)\right|^2, \quad 0\geq x\geq a \tag{4.54}$$

The two expressions can be reconciled by equating the two predictions for the total energy.

$$\int_0^{a/2} U_{Eff}(x,y)dx = \int_0^{a/2} U(x,y)dx \tag{4.55}$$

Symmetry with respect to the center of the slit (x = a/2), allows for integration over only half the region (x = 0 to a/2, instead of x = 0, a). In this case, integration over y is unnecessary; the two expressions have the same variation in the y-direction, determined by the propagation constant, β. From 4.55, a little algebra can provide an expression for the effective dielectric constant.

$$\int_0^{a/2}\left[\frac{1}{2}\mu_0\left|\vec{H}\right|^2 + \frac{1}{2}\varepsilon_{Eff}\left|\vec{E}\right|^2\right]dx = \int_0^{d}\left[\frac{1}{2}\mu_0\left|\vec{H}\right|^2 + \frac{1}{2}\varepsilon_{F}\left|\vec{E}\right|^2\right]dx + \int_d^{a/2}\left[\frac{1}{2}\mu_0\left|\vec{H}\right|^2 + \frac{1}{2}\varepsilon_{D}\left|\vec{E}\right|^2\right]dx \tag{4.56}$$

$$\int_0^{a/2}\mu_0\left|\vec{H}\right|^2 dx + \varepsilon_{Eff}\int_0^{a/2}\left|\vec{E}\right|^2 dx = \int_0^{a/2}\mu_0\left|\vec{H}\right|^2 dx + \varepsilon_{F}\int_0^{d}\left|\vec{E}\right|^2 dx + \varepsilon_{D}\int_d^{a/2}\left|\vec{E}\right|^2 dx \tag{4.57}$$

$$\varepsilon_{Eff} = \frac{\varepsilon_F\int_0^{d}\left|\vec{E}\right|^2 dx + \varepsilon_D\int_d^{a/2}\left|\vec{E}\right|^2 dx}{\int_0^{a/2}\left|\vec{E}\right|^2 dx} \tag{4.58}$$

In our peak shift equation, the change in the slit side dielectric constant (from the uncoated value of ε_D) is used. To determine this, the uncoated value is simply subtracted from 4.58.

$$\delta\varepsilon = \varepsilon_{Eff} - \varepsilon_D = \frac{\varepsilon_F\int_0^{d}\left|\vec{E}\right|^2 dx + \varepsilon_D\int_d^{a/2}\left|\vec{E}\right|^2 dx - \varepsilon_D\int_0^{a/2}\left|\vec{E}\right|^2 dx}{\int_0^{a/2}\left|\vec{E}\right|^2 dx} \tag{4.59}$$

$$\delta\varepsilon = \frac{\varepsilon_F\int_0^{d}\left|\vec{E}\right|^2 dx - \varepsilon_D\int_0^{d}\left|\vec{E}\right|^2 dx}{\int_0^{a/2}\left|\vec{E}\right|^2 dx} \tag{4.60}$$

$$\delta\varepsilon = \left(\varepsilon_F - \varepsilon_D\right)\frac{\int_0^d \left|\vec{E}\right|^2 dx}{\int_0^{a/2} \left|\vec{E}\right|^2 dx} \tag{4.61}$$

In other words, the change in the effective dielectric constant in the slit region is determined by the fraction of the integrated electric field intensity that resides in the analyte coating. If the analyte layer is thin, certainly 4.61 can be approximated to different orders of d.

Predicting the peak shift observed experimentally using the approximate expression in 4.37 and the "exact" - but implicit - method described with 4.35 is a subject of ongoing work. In the following section, preliminary results for one particular aspect of this problem are outlined. This will also give the reader an idea of how the methods discussed at length in this section can be utilized. It will also be shown that the peak shift is not entirely determined by the shift in the quadrupolar resonance, though it is a major component of the overall shift.

4.1.3 Blue Shift Analysis

If dispersion effects are ignored, the denominator in 4.37 is unity. In this case, it is relatively easy to see that the total peak shift should be positive (red shift) after the addition of a chemical film with a higher index than air. This arises because the derivative of the effective index with respect to the dielectric constant of the insulator side is positive and the effective change in the dielectric constant on the insulator side is also positive if you are depositing a high index material on the interface.

To explore this phenomenon, the propagation constant version of the peak shift equation in 4.9 is examined. The real and imaginary components of the propagation constant and the relevant first-derivatives are plotted in Figure 20 for a silver grating structure and in Figure 21

for a gold grating structure. The front of the structure is assumed to be exposed to air and the substrate is assumed to be either amorphous silicon or quartz [6]. For the slit region, a slit width of 100 nm is used.

As expected, the derivative of the effective index with respect to the ambient dielectric constant is positive almost everywhere for both metallic structures (except for a very narrow region near the silver resonant point). This means – from 4.9 – that a blue shift can only be explained by a positive denominator, i.e., by a real propagation constant that increases with wavelength. This occurs near the plasmon resonance for each interface. For the air-metal, quartz-metal, and air-metal slit plots, this resonance occurs near 350 nm in a silver grating and near 500 nm in a gold grating. The high dielectric constant of amorphous silicon shifts the resonance up near 600 nm for silver and 700 nm for gold. This suggests that a silicon substrate would contribute most significantly to an overall blue shift.

The above argument refers to the position of the quadrupolar resonance. If there were no loss in the system, the resonant peak would be infinitely sharp and would be entirely determined by this model. In a system with loss, the wavelength dependence of absorption can also cause an effective peak shift. Consider a system with peak transmission at $\lambda = \lambda_R$ if you ignore loss. Then assume that the loss is decreasing with wavelength; there is more absorption at λ_R-$\Delta\lambda$ and less at λ_R+$\Delta\lambda$. Taking account of loss, that means the transmission is decreased a little at λ_R+$\Delta\lambda$, a little more at λ_R, and drops the most at λ_R-$\Delta\lambda$. The net result is that the peak transmission is shifted to a longer (and lower loss) wavelength. In short, if the imaginary component of the propagation constant is increasing, it contributes a negative (blue) shift. If it is decreasing, the loss causes a positive (red) wavelength shift in the transmission peak. It is somewhat easy to see from the transmission spectrum when this mechanism is significant; the deposition of the analyte layer

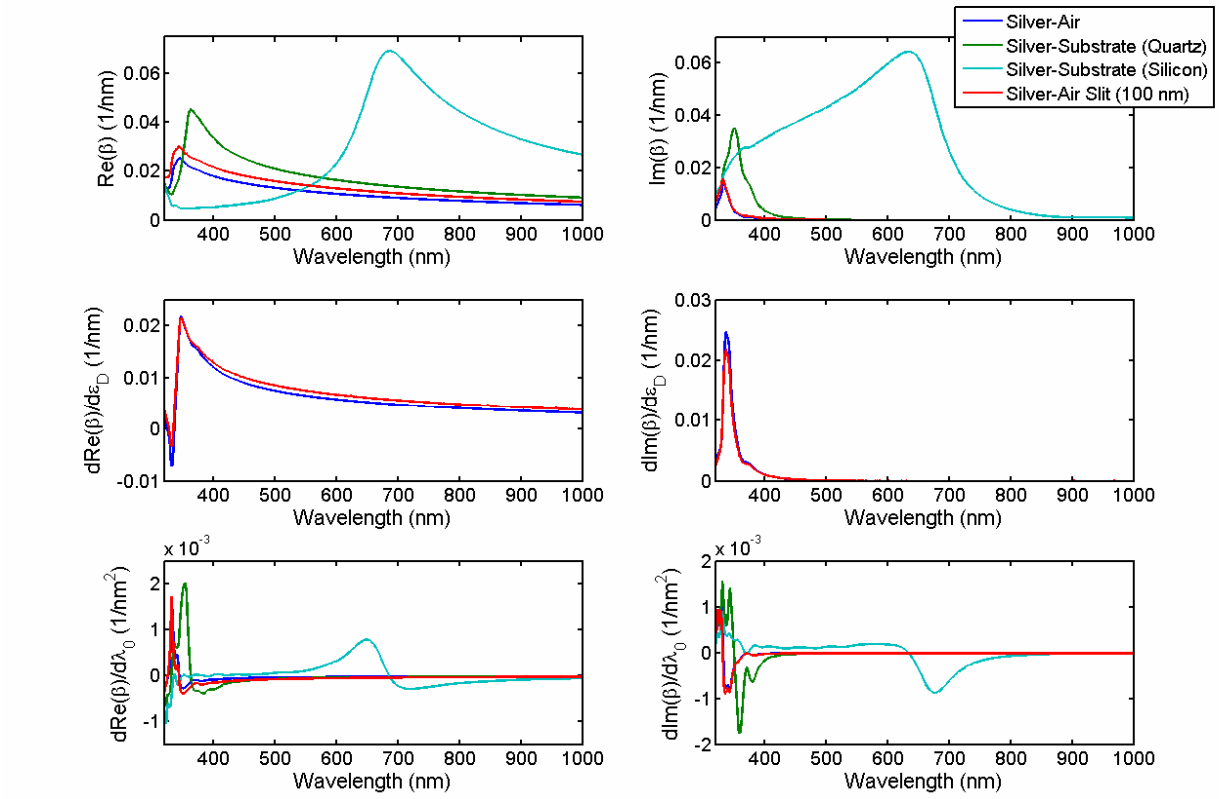

Figure 20: The propagation constant for SP modes associated with different interfaces in a silver slit array structure is plotted over a range of wavelengths.

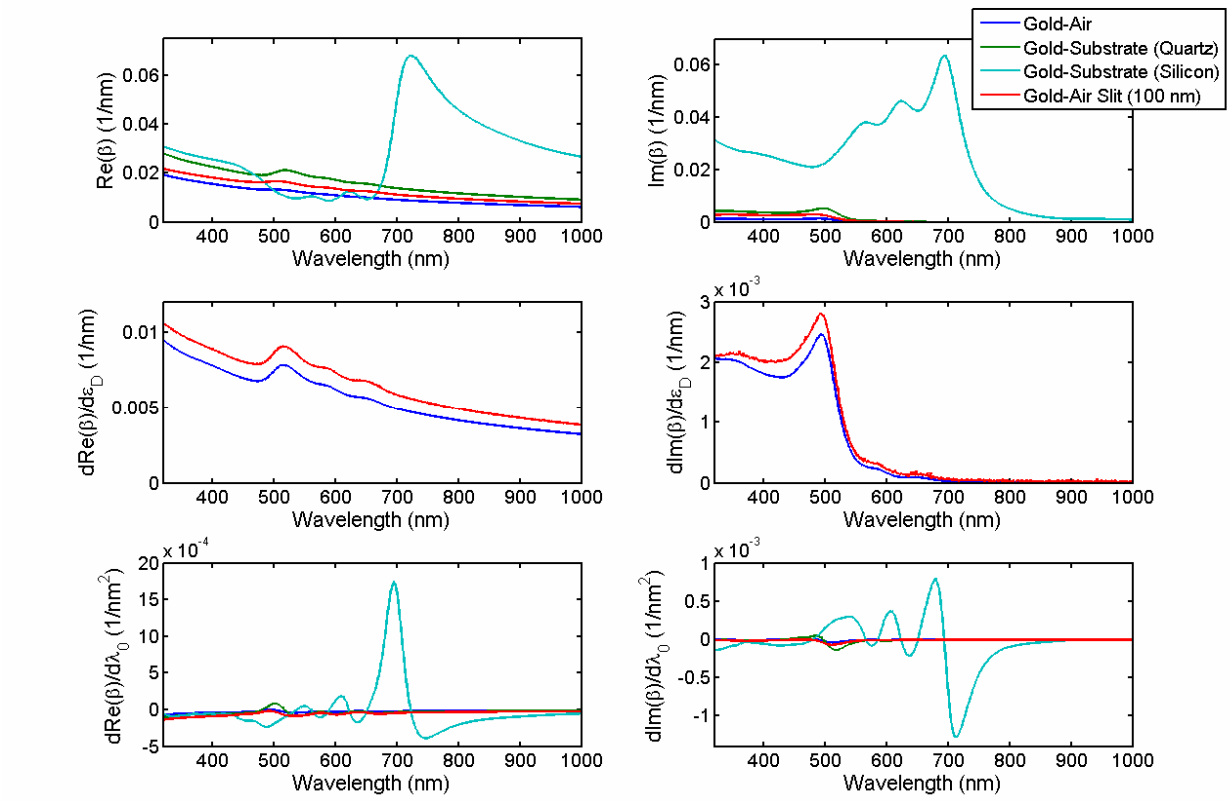

Figure 21: The propagation constant for SP modes associated with different interfaces in a gold slit array structure is plotted over a range of wavelengths.

will cause one side of the transmission peak to drop more than the other side, giving the peak a lop-sided appearance compared to the initial peak. In the context of chemical sensing, we are only interested in the net distortion of the peak after coating with the analyte layer. This effect can be determined by the plot of the change in Im(β) with dielectric constant. Near wavelengths where this function is large, the transmission profile is preferentially attenuated after the chemical deposition. As discussed, this will result in red shift on the left side of the peak and blue shift on the right side of the peak.

Another structure that has been seen to exhibit significant blue shift is a gold grating with a composite silicon substrate, such as a thin layer of amorphous silicon sitting between the gold and the bulk quartz substrate. The above analysis of the bulk silicon substrate has shown that it exhibits very strong dispersion in the wavelength region of interest (~500 – 800 nm). It stands to reason that the effective index of plasmons at this interface will be strongly dependent upon the thickness of the film. Figure 22 shows the variation of the propagation constant over a range of film thickness.

Consider the case of a gold grating on 10 nm of amorphous silicon on a semi-infinite quartz substrate. It is apparent from Figures 21 and 22 that the dispersion associated with the substrate interface will be dominant. While the bulk quartz substrate exhibits anomalous dispersion only in a narrow region around λ_0 = 500 nm, and the bulk silicon substrate exhibits anomalous dispersion only near λ_0 = 670 nm or so, the effect of introducing a "composite" substrate is to smear the anomalous dispersion over a much larger range. For the case of the 10 nm thick silicon, the slope of the effective index is positive from roughly 530 nm to 670 nm. This would seem to enhance the range of wavelengths over which blue shift can occur. It is also apparent from 4.38 that the peak shift is large when the denominator is small, so very weak

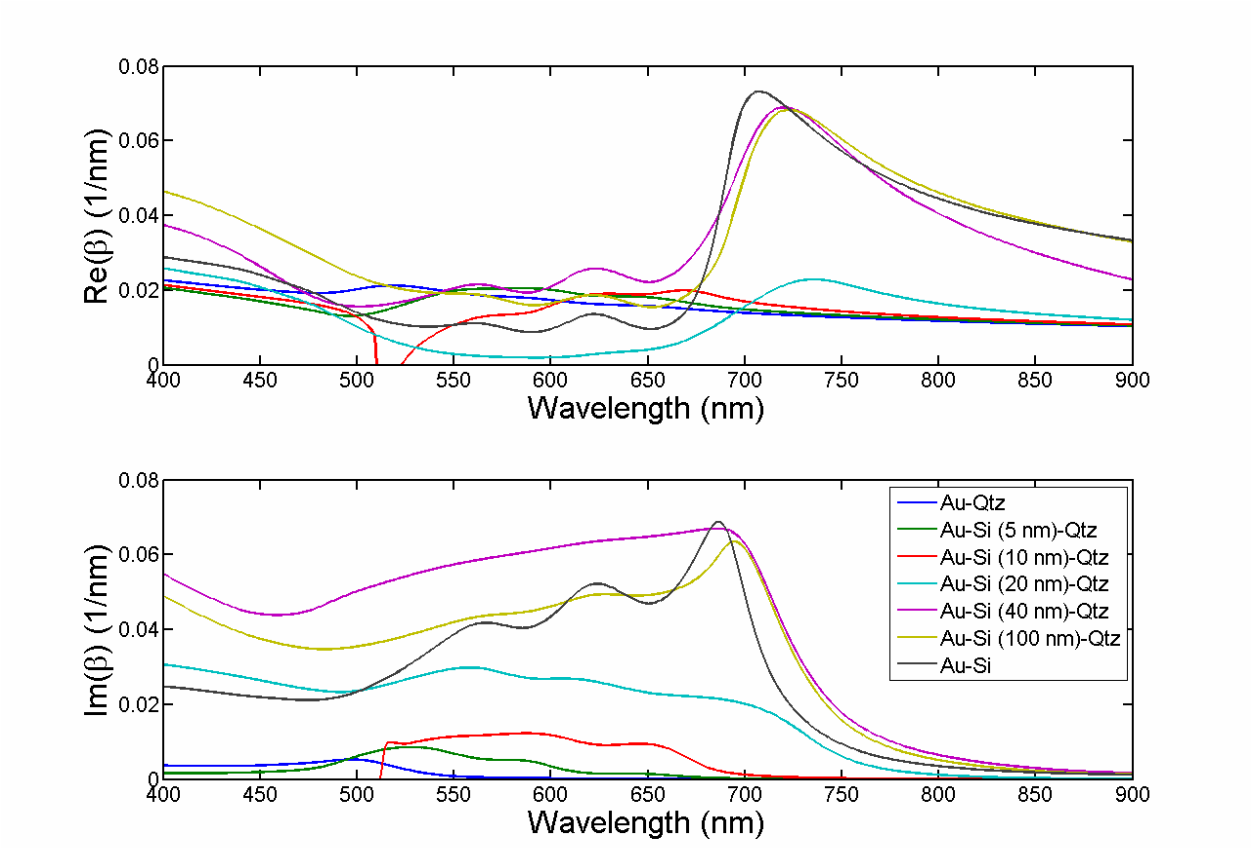

Figure 22: The propagation constant for a layer of amorphous silicon lying between gold and quartz (both assumed to be semi-infinite) plotted over a range of film thickness and λ_0.

positive dispersion would have a more noticeable effect than if it was extremely sharp. From Figure 21, it appears that loss-dependent peak shift would contribute to blue shift only in the 400 to 500 nm region and would contribute to the red shift everywhere else. This seems to suggest that any blue shift observed in this spectral region would be determined primarily by the shift in the resonance due to anomalous dispersion.

4.2 HIGH INDEX SYSTEMS

Shown in Figure 23 is the effective index for an isolated silver slit, calculated using the eigenvalue equation derived in Chapter 2. The effective index is shown over a range of slit widths (from the infinite slit width limit down to 5 nm slit width) and a range of free-space wavelengths (from 337 nm, where $\mathrm{Re}(\varepsilon_M) \approx -1$, to 650 nm). It can be seen that the real part of the effective index reaches a maximum around $\lambda_0 = 350$ nm for all slit widths. In this region, the imaginary part of the effective index tends to be significant.

Consider a nanolens [13],[14] defined by an array of slits which are spaced far enough apart to avoid coupling between slits. The beam shaping effect is achieved by changing the relative phase shift as a function of position: by modulating slit width or film thickness. The phase shift achieved by light of wavelength λ_0 propagating through a film of thickness t; assume the slit has an effective index N.

$$\varphi = kt = \frac{2\pi}{\lambda_0} Nt \tag{4.62}$$

First consider the method where the film thickness is varied from t to t + Δt. Assume this is a relatively small variation. The maximum relative phase shift is as follows.

$$\Delta\varphi_T = kt = \frac{2\pi}{\lambda_0} N\Delta t \tag{4.63}$$

Alternatively, consider the method where the effective index is changed by varying the slit width. The change in N with slit width can be approximated from the plots in Figure 23.

$$\Delta\varphi_N = \frac{2\pi}{\lambda_0} t\Delta N \tag{4.64}$$

These two methods can be compared directly.

$$\frac{\Delta\varphi_N}{\Delta\varphi_T} = \left(\frac{t}{\Delta t}\right)\left(\frac{\Delta N}{N}\right) \tag{4.65}$$

From the plot in Figure 23, it can be seen that effective index is extremely sensitive to changes in the slit width, especially near resonance. Naively, the real part of the effective index could be increased by up to a factor of 14 or 15; suggesting the superiority of the slit-width variation method. However, this is a very rough quantification of the structure's beam shaping ability. The coupling efficiency of a slit may decrease significantly as slit width is decreases, decreasing transmission. Furthermore, the imaginary part of the effective index is also strongly dependent upon the slit width. Regions near resonance – which seem to be the best choice based solely on

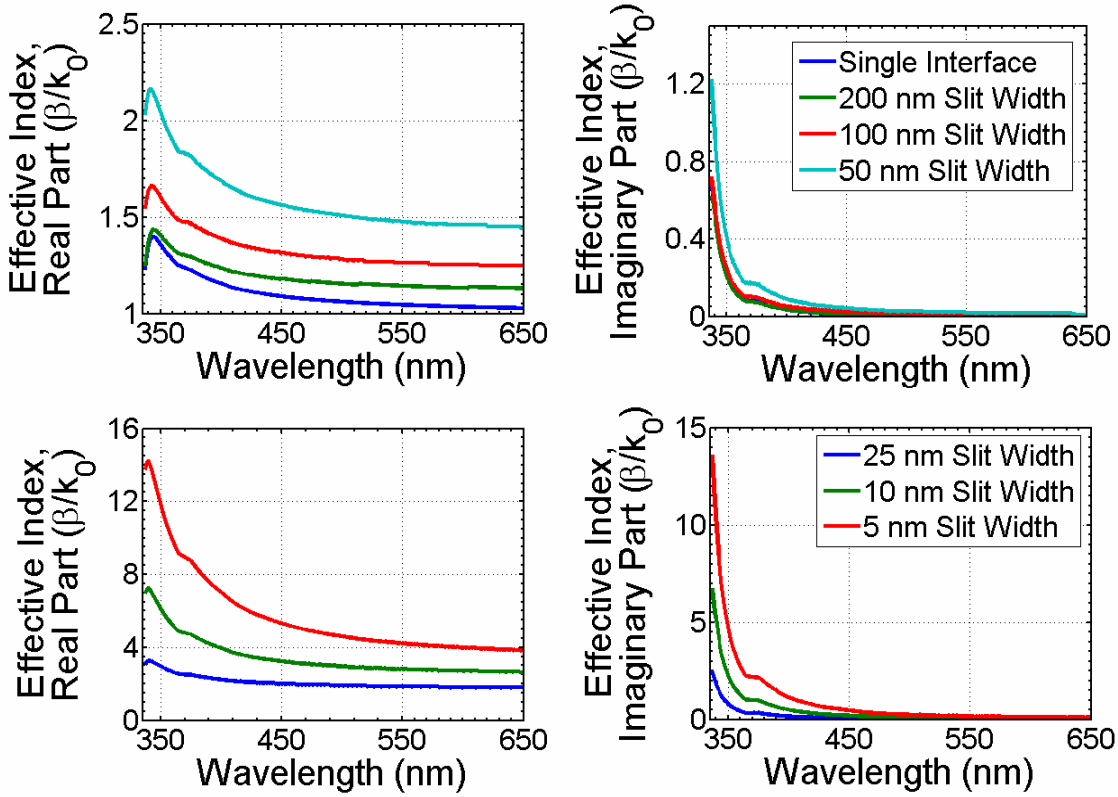

Figure 23: The real and imaginary parts of the effective index of the symmetric SP mode over a wide range of wavelengths (from 650 nm down to resonance). The results for a number of different slit widths are shown, ranging from strong coupling (a = 5 nm) to no coupling (single interface).

the criterion in 4.64 – tend to be very high loss. High index, high loss systems also tend to be highly reflective; how this would affect the coupling efficiency of the slit could also be a major problem. There is no simple analytic solution for the coupling efficiency available, so designing an "optimal" geometry for beam shaping would mostly be a matter of educated-guess-and-check. Furthermore, for the slit width values of 25 nm and below (the lower half of Figure 23), ease and precision of fabrication may become a significant issue.

All in all, if the design is limited to low loss frequencies (sufficiently far away from resonance) and reasonably large values of slit width (50 nm and above), slit-width variation is not – in principle – much better than thickness variation. If the slit width is more than 50 nm and the wavelength more than 400 nm or so, the effective index can be increased by a factor of two; by the same token, increasing the thickness by a factor two is not at all unreasonable.

4.3 METALLIC-DIELECTRIC STACKS

In the previous two applications, we focused specifically on systems which utilized propagating surface plasmons. This section will explore a slightly different system – a distributed Bragg reflector (DBR) with metallic layers. A "traditional" DBR is a type of one-dimensional photonic crystal, which consists of a layer of periodic, alternating stacks of different dielectric materials. The analysis of such periodic structures is presented in many modern photonics textbooks; the end result of this analysis – depending on the periodicity and dielectric contrast of the stacks – is that a photonic bandgap can form. In other words, there will be a range of frequencies that cannot propagate. This type of system can be utilized as an optical bandstop filter.

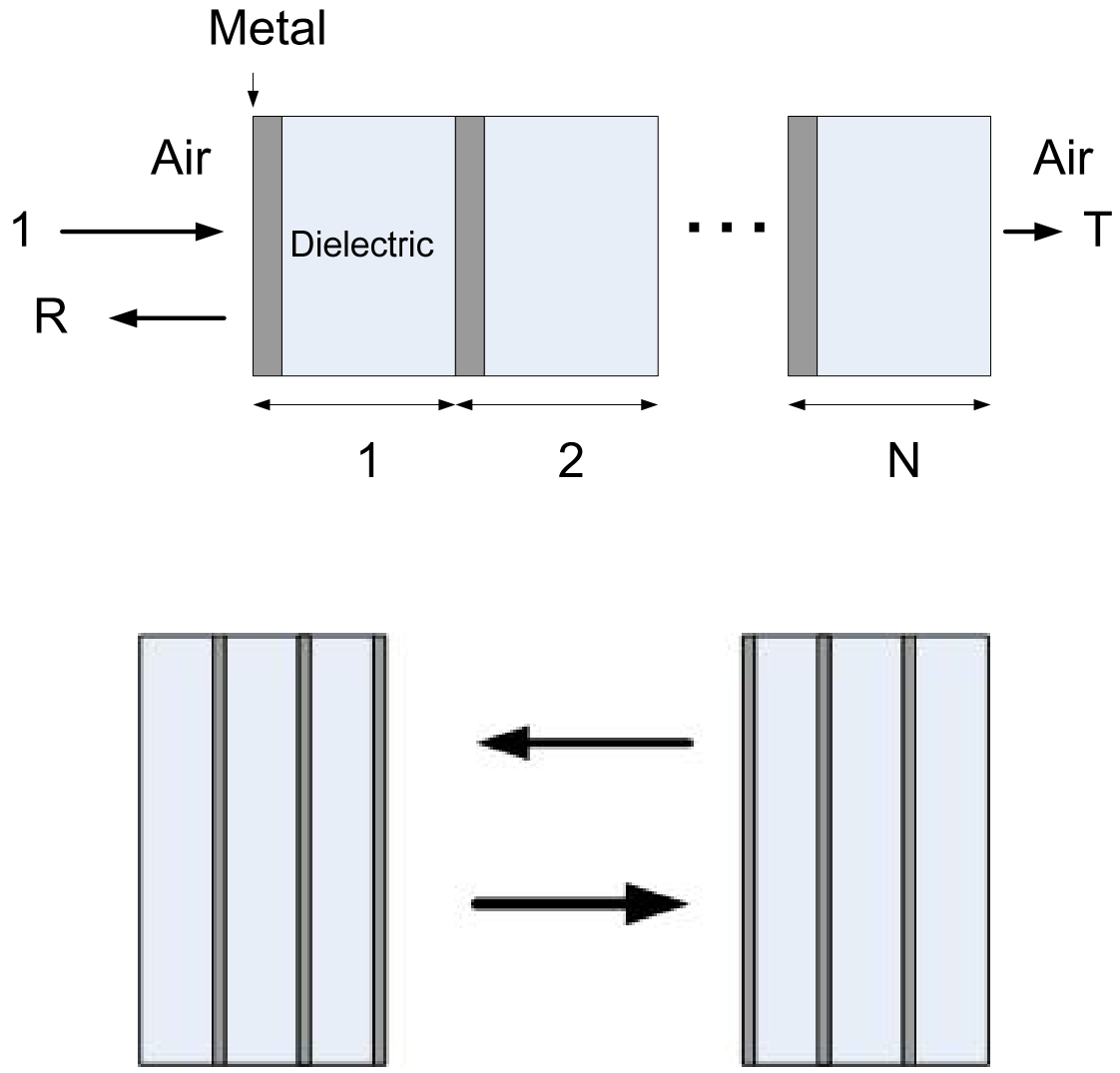

Figure 24: (Top) A sketch of the metal-dielectric alternating structure described in the text (N period). R and T are the fractions of the normalized incident power which are reflected off the in the structure and will be strongly reflected.

At first glance, there should not be anything essentially different with a stack that dielectric DBRs already exhibit very high reflectivity in the band gap region. It seems like all the metal could provide is absorption, not a particularly attractive contribution. But it turns out that metallo-dielectric DBRs (MD-DBR) can exhibit some interesting properties.front interface and transmitted through the stack. (Bottom) A resonant cavity formed by two such structures; the gap is filled with air or some other dielectric.

The analysis for a periodic structure can be complicated, but the end result is relatively simple. Utilizing the transfer-matrix formalism (discussed in a different context in Appendix B), the boundary condition information can be concisely expressed in matrix form (modified slightly from Yariv and Yeh [1]).

$$\begin{pmatrix} 1 & 1 \\ \gamma_M & -\gamma_M \end{pmatrix}\begin{pmatrix} h_n^{M,+} \\ h_n^{M,-} \end{pmatrix} = \begin{pmatrix} e^{-ik_D b} & e^{ik_D b} \\ ik_D e^{-ik_D b} & -ik_D e^{ik_D b} \end{pmatrix}\begin{pmatrix} h_{n+1}^{D,+} \\ h_{n+1}^{D,-} \end{pmatrix} \tag{4.66}$$

$$\begin{pmatrix} e^{\gamma_M a} & e^{-\gamma_M a} \\ \gamma_M e^{\gamma_M a} & -\gamma_M e^{-\gamma_M a} \end{pmatrix}\begin{pmatrix} h_n^{M,+} \\ h_n^{M,-} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ ik_D & -ik_D \end{pmatrix}\begin{pmatrix} h_n^{D,+} \\ h_n^{D,-} \end{pmatrix} \tag{4.67}$$

Here $h_n^{M,\pm}$ refers to the 'forward' (+z) and 'backward' (-z) decaying components in the metal (with similar notation for the forward and backward propagating components in the dielectric), a is the metal thickness, and b is the dielectric thickness. For simplicity, assume normal incidence. The wave vector in the dielectric and the decay constant in the metal are given as follows.

$$\gamma_M = \sqrt{-\varepsilon_M}\, k_0 \tag{4.68}$$

$$k_D = \sqrt{\varepsilon_D}\, k_0 \tag{4.69}$$

The matrix relations provide a nice recursion relationship between the nth dielectric layer and the (n+1)th dielectric layer.

$$\begin{pmatrix} h_{n+1}^{D,+} \\ h_{n+1}^{D,-} \end{pmatrix} = \begin{pmatrix} e^{-ik_D b} & e^{ik_D b} \\ ik_D e^{-ik_D b} & -ik_D e^{ik_D b} \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ \gamma_M & -\gamma_M \end{pmatrix} \begin{pmatrix} e^{\gamma_M a} & e^{-\gamma_M a} \\ \gamma_M e^{\gamma_M a} & -\gamma_M e^{-\gamma_M a} \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ ik_D & -ik_D \end{pmatrix} \begin{pmatrix} h_n^{D,+} \\ h_n^{D,-} \end{pmatrix} \tag{4.70}$$

This can certainly be simplified, but if the wavelength, metal thickness, and dielectric thickness are given, this expression can be easily calculated computationally. For notational simplicity, re-write the combination of known transfer matrices above as a single matrix.

$$\begin{pmatrix} h_{n+1}^{D,+} \\ h_{n+1}^{D,-} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} h_n^{D,+} \\ h_n^{D,-} \end{pmatrix} \tag{4.71}$$

This also means that the input at the first layer can easily be related to output at the last (Nth) layer.

$$\begin{pmatrix} h_N^{D,+} \\ h_N^{D,-} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^N \begin{pmatrix} h_0^{D,+} \\ h_0^{D,-} \end{pmatrix} \equiv \begin{pmatrix} A_N & B_N \\ C_N & D_N \end{pmatrix} \begin{pmatrix} h_0^{D,+} \\ h_0^{D,-} \end{pmatrix} \tag{4.72}$$

If the incident light is normalized, this can be written in terms of the reflected and transmitted light.

$$\begin{pmatrix} t \\ 0 \end{pmatrix} = \begin{pmatrix} A_N & B_N \\ C_N & D_N \end{pmatrix} \begin{pmatrix} 1 \\ r \end{pmatrix} \tag{4.73}$$

A few steps of algebra provides a simple solution for the reflection, transmission, and loss of the system.

$$R = |r|^2 = \frac{|C_N|^2}{|D_N|^2} \tag{4.74}$$

$$T = |t|^2 = \frac{|A_N D_N - B_N C_N|^2}{|D_N|^2} \tag{4.75}$$

$$L = 1 - R - T = \frac{|D_N|^2 - |C_N|^2 - |A_N D_N - B_N C_N|^2}{|D_N|^2} \tag{4.76}$$

In lossless systems, conservation of energy requires that the determinant of the transfer matrix (for a given layer or overall) is equal to unity [1]. This relationship would simplify the expression for T. In our system, energy is lost in the metal layers and the transfer matrix is not unimodular.

Consider the design of a metal-dielectric Bragg stack for frequencies in the near-UV region of the spectrum - for example, at λ_0 = 325 nm (one standard wavelength for HeCd lasers). In order to properly exploit the interesting properties of the metal (negative dielectric constant), a metal with a plasmon frequency above the frequency of interest should be selected; otherwise it is no different than using a simple dielectric (except it would be lossier). Metallic aluminum has a plasmon frequency of $\hbar\omega_P \approx 13eV$ and a dielectric constant of -15.9 + 2.4i at λ_0 = 325 nm, making it a good candidate [6]. For the sake of argument, consider a glass with index 1.5 for the dielectric (approximately true for SiO_2 at 325 nm). There are three additional design parameters which can be utilized: metal layer thickness, glass layer thickness, and the number of layers. In a dielectric-dielectric DBR, more layers is almost always better; if no loss is assumed in the system, additional layers simply gives the system a sharper photonic band gap and decreases Fabry-Perot cavity effects between the front and back faces. In fact, the ideal dielectric-dielectric system is optimal when the number of layers goes to infinity. In a metallic-dielectric system, the

light loses energy every time it passes through a metallic layer; so it is easy to see there is a bit of a trade-off regarding the number of layers. Figure 25 shows the transmission, reflection, and loss for an N-layer aluminum-glass stack at $\lambda_0 = 325$ nm for several different values of N and a broad range of glass thickness. The overall metal thickness should not be large compared to the penetration depth; a plane wave propagating into bulk metallic aluminum (based on the dielectric constant cited above) will lose roughly 63% of its power after propagating only 6.5 nm – though we hope that the effect of periodicity will somehow lessen the loss. To this effect, a metal thickness of 10 nm is selected.

The transmission spectrum has a few features that could be of interest. The results are compared to the result for a semi-infinite metallic aluminum mirror in Figure 25. It can be seen that at certain values of glass thickness, both the reflectivity and loss are improved over the bulk metallic mirror (reflectivity is larger and loss is smaller than the bulk case).

In the study of resonant cavities, the Q-factor is used as a quantitative measure of the quality of the cavity. Physically it is related to how well the cavity holds energy and is related to the line-width of the resulting resonant modes (modes in a cavity with a very high Q factor will have very sharp resonant modes). It is defined as follows [3].

$$Q \equiv \omega_0 \frac{(Total \quad Energy \quad in \quad Cavity)}{(Loss)} \tag{4.77}$$

Consider S_0 power per unit area bouncing back and forth in a cavity over one period. The total loss for the cavity will be the sum of the transmission and the absorption (everything except the reflected power); during a full round trip, the light hits two interfaces, causing an overall loss of $2S_0(1\text{-}R)$. The total *energy* in the cavity will be equal to the total power passing a particular point

in a given period, S_0T. Based on the definition above, the Q factor for a resonant cavity of this nature can be approximated.

$$Q \equiv \omega_0 \frac{S_0 T}{2S_0(1-R)} = \frac{\pi}{1-R} \tag{4.78}$$

This geometry is sketched in the bottom part of Figure 24 and the results are plotted in Figure 26. A quality factor of 100 is not extremely impressive; this result serves to illustrate the applicability of the DBR idea to metallic systems and to show that increasing the number of layers does improve the optical properties of the system, despite the expectation that more metal layers should imply more loss.

The results shown here focused on relatively thin metallic layers (less than the penetration depth). There have been studies done on MD-DBRs with relatively thick metallic layers. A naïve look at the system would suggest that the maximum possible transmission is limited by the power lost in each layer. Work by Scalora et al [26] has illustrated that this is not the case. Among other things, they show that a 20 layer stack of silver and MgF_2 with a overall silver thickness of 200 nm (10 nm per layer) can have significantly higher peak transmission than a 200 nm thick silver film (almost 40% peak transmission for the MD-DBR around $\lambda_0 = 400$ nm, "about seven orders of magnitude better transmission than for the single 200 nm silver film case" [26]). This also shows that the system designed above as a demonstrative tool can be improved immensely with more careful design (for example, with the selection of a higher-index dielectric – such as MgF_2).

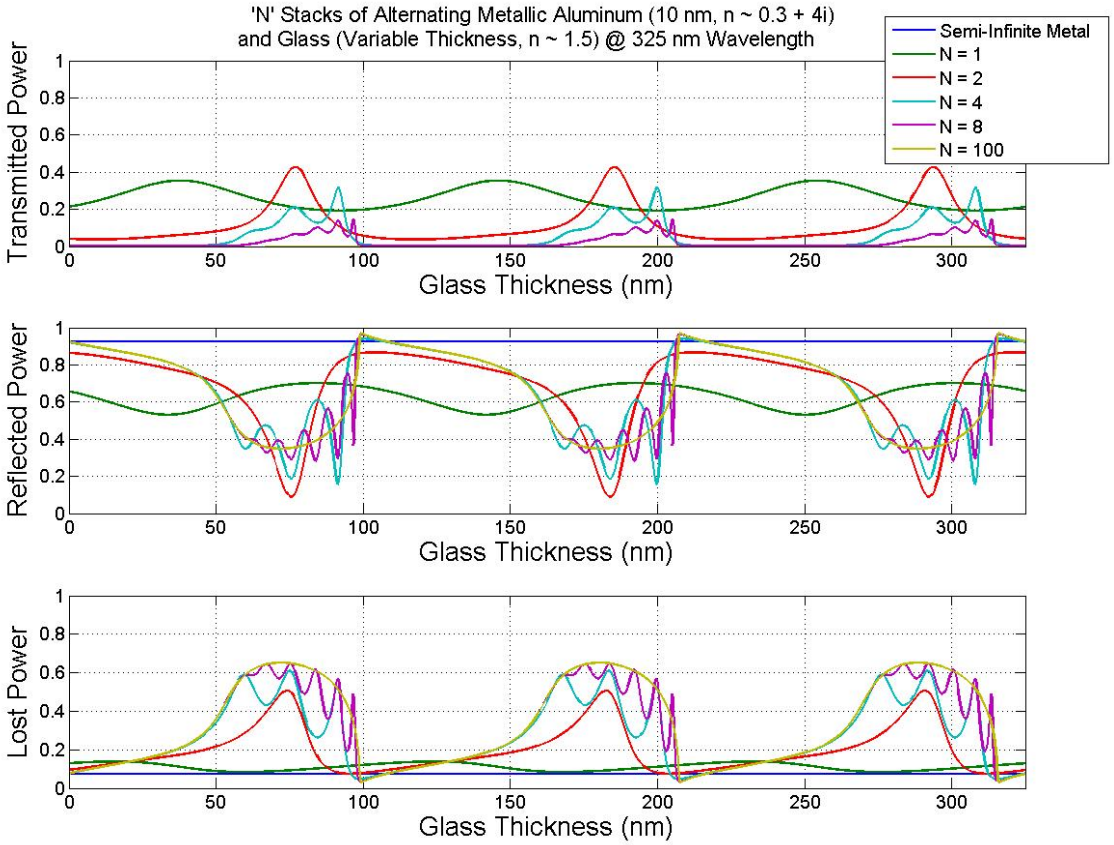

Figure 25: Transmission, reflection and loss from light propagating in an aluminum-glass stack in the near UV, the glass layer thickness and the number of stacks are varied to show where the geometry is optimal.

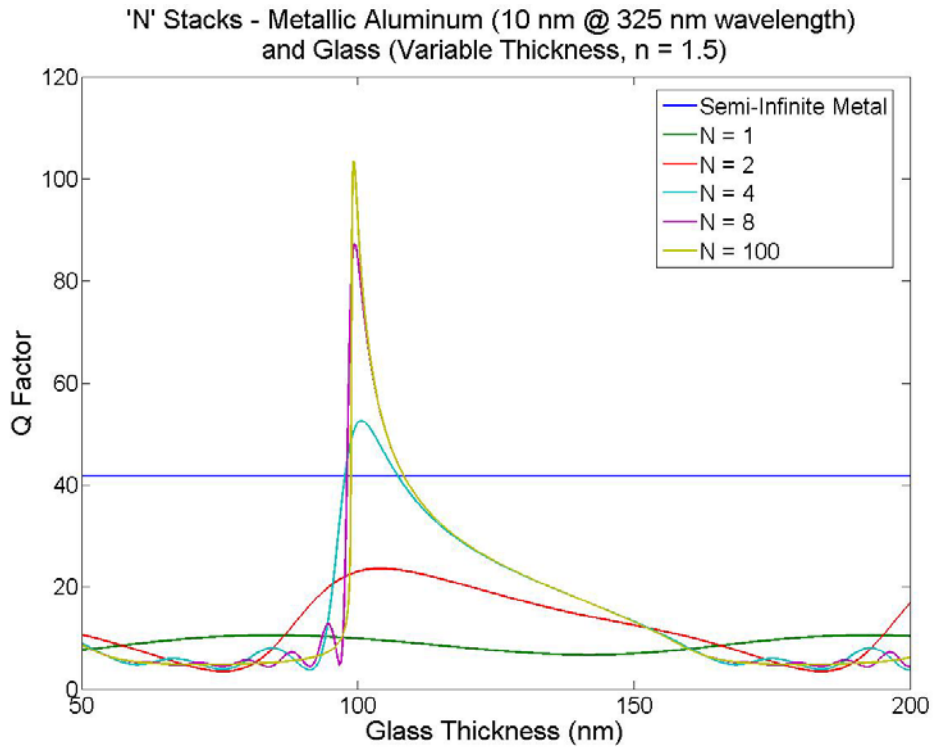

Figure 26: Q factor of a resonant cavity formed by two N-period aluminum-glass sticks filled with air. The glass thickness and the number of layers is varied.

5.0 CONCLUSION

It may be hard to believe that so much can be written about a structure as simple as a slit array in a metallic film. It may be equally suprising that anything new or interesting could be said about such a simple structure. Arguably, apertures have been studied longer than any other optical device [10]. The novelty and recent research interest in this topic lies in the relatively newfound ability to fabricate structures with dimensions on the order of optical wavelengths in conjunction with the interesting properties of metals at optical frequencies.

Some works have studied the same structure modeled as a perfect electric conductor [8]. This approach is helpful because it lends itself to relatively simple analytic solutions. It was shown in Chapter 2 and 3 that many of the interesting optical properties of metallic nano-apertures arise from the negative permittivity of the metal at certain frequencies. It was shown in Chapter 2 that if the magnitude of the dielectric constant is very large (essentially infinite) the mode and effective index properties are independent of the complex phase. The optics will be essentially the same whether the waveguide is a PEC or a metal with a strongly negative index of refraction. This means that the perfect conductor assumption loses something, especially if the excitation frequency is near enough to the plasmon frequency to make the dielectric constant real and negative, but not large enough to be approximated as infinite. In other words, the field penetration into the system is not negligible. It has been suggested throughout this work that the reversal of power flow on the metal side may be an important aspect of the funneling

phenomenon; if so, this property would almost certainly be lost if the negative dielectric constant is not properly taken into consideration.

In more recent and rigorous works, the metal is modeled with a negative, but completely real dielectric constant. This has the advantage of at least partially capturing the plasmonic nature of the TM_0 mode, but misses some very important aspects of the problem. It was shown in Chapter 2 that the TM_0 mode is distinguished from higher order modes because its loss is determined only by the complex dielectric constant of the metal. If the symmetric SP mode is as important as expected, the limitations of this mode (in terms of propagation length, coupling efficiency, etc.) will be strongly related to the complex nature of the metallic dielectric constant. Furthermore, the imaginary part of the dielectric constant is relatively large near the resonant frequency, as shown in Figure 1. For silver or gold, this means that the optical properties of the structure in the blue and near UV regions of the spectrum are strongly dependent upon the complex nature of the dielectric constant.

In Chapter 4, the benefits of this structure in the area of chemical sensing were emphasized. Many sensing applications essentially utilize single interface systems. It has been discussed in this work that the sensing capability of the structure is primarily influenced by the slit region. In the single interface case, the resonance conditions are very sensitive to surface treatment because of the natural surface confinement of the SP wave. In the slit region, the total propagating power is confined over an even shorter range – as limited by the slit width – while the wave profile is still strongest near the interfaces.

The chemical sensing properties of the metallic nano-slit array have been studied experimentally [27], so the above prediction of high sensitivity has been justified (at least qualitatively). In terms of future work, the ultimate goal should be to establish a complete model

for light coupling and transmission in an array of nano-slits. This problem can be divided up into three parts: (1) light propagation in a metallic nanoslit, (2) coupling in and out of the slit, and (3) the role of resonant modes in the structure. The first topic was explored extensively in this thesis; only preliminary results were discussed for the second two topics. One would think that traditional coupled mode theory should adequately describe coupling into and out of the slit. But there are some subtle issues regarding the existence of bound modes propagating along the front interface, which may make the problem more difficult. From that perspective, it could be argued that the latter two aspects of the problem are actually different ways of looking at the same thing, particularly if one is only interested in peaks and dips in the transmission spectrum. The discussion of coupling and transmission in Chapter 2 mostly utilized numerical simulations; while this approach might be useful in gaining a qualitative understanding of the problem, a more analytic solution would eventually be preferable for design and optimization purposes.

APPENDIX A

MODELING METALLIC SYSTEMS WITH FDTD

For physically interesting systems, solving for the scattered fields directly from Maxwell's equations can be difficult. Often, approximating or isolating certain parts of the system can obscure some of the more subtle aspects of the physical problem. In such cases it is easier to study the system computationally. If the system is correctly modeled, the accuracy is then limited only to the computational power devoted to the task. The purpose of this appendix is to sketch out a technique for simulating a metallic system which cannot be accurately modeled as a PEC. For a more in-depth and thorough discussion of FDTD as a general tool,

One such method for studying the interaction between light (modeled using Maxwell's equations) and material structures (modeled with a dielectric constant which is a function of position) is the finite-difference time domain method (FDTD). It begins by utilizing the curl equations for the electric and magnetic fields.

$$\nabla \times \vec{H}(\vec{r},t) = \frac{1}{c}\frac{\partial \vec{D}}{\partial t}(\vec{r},t) \tag{A.1}$$

$$\nabla \times \vec{E}(\vec{r},t) = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t}(\vec{r},t) \tag{A.2}$$

And the constitutive relations,

$$\vec{D}(\vec{r},t) = \varepsilon(\vec{r})\vec{E}(\vec{r},t) \tag{A.3}$$

$$\vec{B}(\vec{r},t) = \vec{H}(\vec{r},t) \tag{A.4}$$

In general, the method presented here is only one way of discretizing Maxwell's equations. Of course, any particular scheme is not guaranteed to be accurate or stable. We will utilize the 2D Yee mesh prescribed in many major works on the subject [28],[29]. The mesh utilized for this finite difference scheme is shown in Figure 27. The electric field grids are offset from the magnetic field grid by a half-step. They are also offset by a half-step on the time axis (meaning, of course, that the nth iteration of the electric field is calculated after the nth, but before the (n+1)th iteration of the magnetic field.

For a simple dielectric system, the update equations are relatively straightforward. Metallic systems (or any other system where the optical properties are strongly dependent on frequency) are somewhat more difficult. We will outline the results derived in Sullivan's text [28].

$$H_z^{(n)}[ii, jj] = H_z^{(n-1)}[ii, jj] + c\Delta t\left(\frac{E_x^{(n-1/2)}[ii, jj] - E_x^{(n-1/2)}[ii, jj-1]}{\Delta y} - \frac{E_y^{(n-1/2)}[ii, jj] - E_y^{(n-1/2)}[ii-1, jj]}{\Delta x}\right) \tag{A.5}$$

$$D_x^{(n+1/2)}[ii, jj] = D_x^{(n-1/2)}[ii, jj] + \left(\frac{c\Delta t}{\Delta y}\right)\left(H_z^{(n)}[ii, jj+1] - H_z^{(n)}[ii, jj]\right) \tag{A.6}$$

$$D_y^{(n+1/2)}[ii, jj] = D_y^{(n-1/2)}[ii, jj] + \left(\frac{c\Delta t}{\Delta x}\right)\left(H_z^{(n)}[ii, jj] - H_z^{(n)}[ii+1, jj]\right) \tag{A.7}$$

$$E_x^{(n+1/2)}[ii, jj] = D_x^{(n+1/2)}[ii, jj] / \varepsilon[ii, jj] \tag{A.8}$$

$$E_y^{(n+1/2)}[ii, jj] = D_y^{(n+1/2)}[ii, jj] / \varepsilon[ii, jj] \tag{A.9}$$

As explained in Chapter 1, the dielectric constant of a nearly-free electron metal can be written as follows.

$$\varepsilon(\omega) = 1 - \frac{\omega_P^2}{\omega(\omega + i\nu)} \tag{A.10}$$

The two parameters - ω_P and ν - are material properties which are related to the resonance position and linewidth of the dielectric response of the material. By utilizing the constitutive relation for the dielectric flux density and performing a z-transform (see Sullivan for details [28]), a discrete time-domain relationship between the dielectric between the flux density and the field can be derived (which will be true for any frequency accurately modeled by the expression above) in terms of two auxiliary fields.

$$E_x^{(n+1/2)}[ii, jj] = D_x^{(n+1/2)}[ii, jj] - S_x^{(n-1/2)}[ii, jj] \tag{A.11}$$

$$S_x^{(n+1/2)}[ii, jj] = (1 + e^{-\nu\Delta t}) S_x^{(n-1/2)}[ii, jj] - e^{-\nu\Delta t} S_x^{(n-3/2)}[ii, jj] + \frac{\omega^2 \Delta t}{\nu} E_x^{(n+1/2)}[ii, jj] \tag{A.12}$$

$$S_x^{(n-3/2)}[ii, jj] = S_x^{(n-1/2)}[ii, jj] \tag{A.13}$$

$$S_x^{(n-1/2)}[ii, jj] = S_x^{(n+1/2)}[ii, jj] \tag{A.14}$$

$$E_y^{(n+1/2)}[ii, jj] = D_y^{(n+1/2)}[ii, jj] - S_y^{(n-1/2)}[ii, jj] \tag{A.15}$$

$$S_y^{(n+1/2)}[ii, jj] = (1 + e^{-\nu\Delta t}) S_y^{(n-1/2)}[ii, jj] - e^{-\nu\Delta t} S_y^{(n-3/2)}[ii, jj] + \frac{\omega^2 \Delta t}{\nu} E_y^{(n+1/2)}[ii, jj] \tag{A.16}$$

$$S_y^{(n-3/2)}[ii, jj] = S_y^{(n-1/2)}[ii, jj] \tag{A.17}$$

$$S_y^{(n-1/2)}[ii, jj] = S_y^{(n+1/2)}[ii, jj] \tag{A.18}$$

One disadvantage of this method is obvious. For the simple dielectric expressions in A.5-A.9, five fields are sufficient to model the system. For the metallic model shown above, four auxiliary fields are added (because both S_x and S_y must be saved in memory for two time steps). The additional fields will almost double the memory requirements. Fortunately this version of the update equation can be localized to the metal itself, which may be a small piece of the overall problem space. Note that these auxiliary fields should not be confused with the pointing vector, which is also represented with S in previous chapters; this Appendix is the only place in the thesis which utilizes these auxiliary fields, so there was no real need to switch from Sullivan's notation.

The slit structure which is of primary interest in this work is very well-suited to this type of finite difference scheme, primarily because it is an inherently rectangular shape. In this particular case, the boundary conditions are implicitly satisfied by the use of the electric flux density in the update equations. A more complicated structure, particularly one containing curves and sharp points, would be significantly more difficult to model.

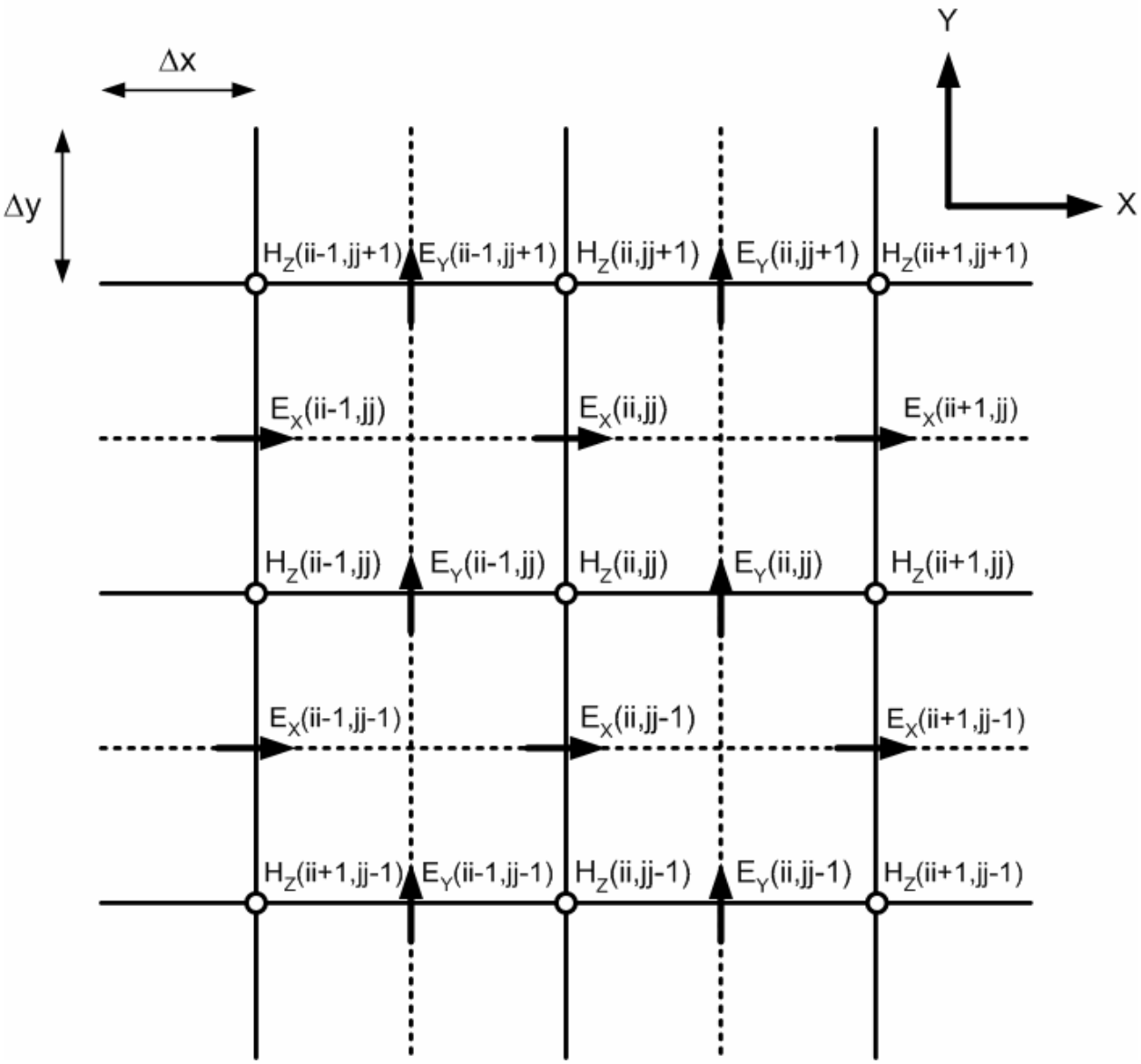

Figure 27: The Yee spatial mesh utilized for calculating the finite-difference equations relating the E_x, E_y, and H_z fields.

APPENDIX B

THE TRANSFER MATRIX METHOD FOR PLASMONIC STRUCTURES

In the main text the eigenmode equations for a coated and uncoated metallic slit were derived. In general, it may be useful to have a more general formulation for deriving the resonant modes of a more complicated structure. For example, consider a simple silver slit structure, coated first with a thin layer of gold, then coated with another thin layer of analyte. Certainly approximations can be made if the coatings are sufficiently thin. But with computational tools like MATLAB or Mathematica, it is relatively easy to numerically evaluate and solve relatively complex matrix equations, so approximation may not be necessary.

An arbitrarily large stack of materials whose optical properties can be adequately characterized through a complex dielectric constant will be considered. A sketch of one layer in the geometry is shown in Figure 28. Define the fields inside the Nth layer in terms of positive and negative exponentials.

$$H_z^{(N)} = e^{i\beta y}\left[H_+^{(N)} e^{\gamma_N (x-x_N)} + H_-^{(N)} e^{-\gamma_N (x-x_{N-1})}\right] \tag{B.1}$$

$$H_z^{(N+1)} = e^{i\beta y}\left[H_+^{(N+1)} e^{\gamma_{N+1} (x-x_{N+1})} + H_-^{(N+1)} e^{-\gamma_{N+1} (x-x_N)}\right] \tag{B.2}$$

In the Nth layer,

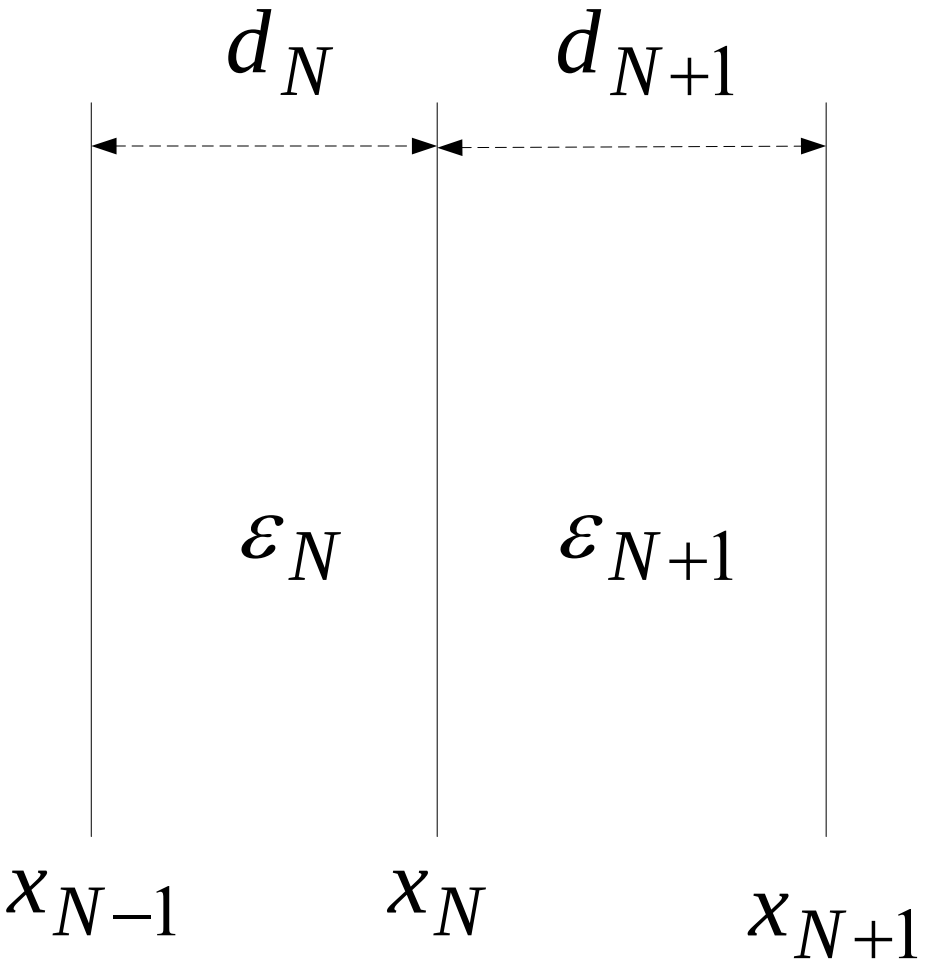

Figure 28: A sketch of two layers (the Nth and (N+1)th layer) in a stack of optical materials. The layer is completely characterized by its dielectric constant and its thickness (d_N).

$$\gamma_N(\beta) = \sqrt{\beta^2 - \varepsilon_N k_0^2} \tag{B.3}$$

Continuity of the magnetic field requires the following

$$H_+^{(N)} + H_-^{(N)} e^{-\gamma_N (x_N - x_{N-1})} = H_+^{(N+1)} e^{\gamma_{N+1}(x_N - x_{N+1})} + H_-^{(N+1)} \tag{B.4}$$

The electric field components follow directly from the magnetic field. For the purpose of deriving the eigenvalue relationship, only the component tangential to the interface(s) is needed.

$$E_y^{(N)} = \frac{1}{i\omega\varepsilon_N} \frac{\partial H_z^{(N)}}{\partial x} \tag{B.5}$$

$$E_y^{(N)} = \frac{\gamma_N}{i\omega\varepsilon_N} e^{i\beta y} \left[H_+^{(N)} e^{\gamma_N (x - x_N)} - H_-^{(N)} e^{-\gamma_N (x - x_{N-1})} \right] \tag{B.6}$$

$$E_y^{(N+1)} = \frac{\gamma_{N+1}}{i\omega\varepsilon_{N+1}} e^{i\beta y} \left[H_+^{(N+1)} e^{\gamma_{N+1}(x - x_{N+1})} - H_-^{(N+1)} e^{-\gamma_{N+1}(x - x_N)} \right] \tag{B.7}$$

At x = x(N), continuity of the y-component of the electric field places an additional constraint.

$$\frac{\gamma_N}{\varepsilon_N} \left[H_+^{(N)} - H_-^{(N)} e^{-\gamma_N (x_N - x_{N-1})} \right] = \frac{\gamma_{N+1}}{\varepsilon_{N+1}} \left[H_+^{(N+1)} e^{\gamma_{N+1}(x_N - x_{N+1})} - H_-^{(N+1)} \right] \tag{B.8}$$

Define

$$d_N = x_N - x_{N-1} \tag{B.9}$$

Therefore, the two boundary conditions are as follows.

$$H_+^{(N)} + H_-^{(N)} e^{-\gamma_N d_N} = H_+^{(N+1)} e^{-\gamma_{N+1} d_{N+1}} + H_-^{(N+1)} \tag{B.10}$$

$$H_+^{(N)} - H_-^{(N)} e^{-\gamma_N d_N} = \frac{\gamma_{N+1}}{\varepsilon_{N+1}} \frac{\varepsilon_N}{\gamma_N} \left(H_+^{(N+1)} e^{-\gamma_{N+1} d_{N+1}} - H_-^{(N+1)} \right) \tag{B.11}$$

To solve for the Nth + and – magnetic components, add B.10 and B.11 together.

$$H_+^{(N)} = \frac{1}{2} \left(1 + \frac{\gamma_{N+1}}{\varepsilon_{N+1}} \frac{\varepsilon_N}{\gamma_N} \right) e^{-\gamma_{N+1} d_{N+1}} H_+^{(N+1)} + \frac{1}{2} \left(1 - \frac{\gamma_{N+1}}{\varepsilon_{N+1}} \frac{\varepsilon_N}{\gamma_N} \right) H_-^{(N+1)} \tag{B.12}$$

Then subtract B.11 from B.10.

$$H_-^{(N)} = \frac{1}{2} \left(1 - \frac{\gamma_{N+1}}{\varepsilon_{N+1}} \frac{\varepsilon_N}{\gamma_N} \right) e^{\gamma_N d_N - \gamma_{N+1} d_{N+1}} H_+^{(N+1)} + \frac{1}{2} \left(1 + \frac{\gamma_{N+1}}{\varepsilon_{N+1}} \frac{\varepsilon_N}{\gamma_N} \right) e^{\gamma_N d_N} H_-^{(N+1)} \tag{B.13}$$

The expression in B.13 is a bit complicated. So to simplify a bit, define the following parameter.

$$\alpha_N = \frac{\gamma_{N+1}}{\varepsilon_{N+1}} \frac{\varepsilon_N}{\gamma_N} \tag{B.14}$$

Making the substitution in B.14, we write the two requirements in matrix form.

$$\begin{pmatrix} H_+^{(N)} \\ H_-^{(N)} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} (1+\alpha_N) e^{-\gamma_{N+1} d_{N+1}} & 1-\alpha_N \\ (1-\alpha_N) e^{\gamma_N d_N - \gamma_{N+1} d_{N+1}} & (1+\alpha_N) e^{\gamma_N d_N} \end{pmatrix} \begin{pmatrix} H_+^{(N+1)} \\ H_-^{(N+1)} \end{pmatrix} \tag{B.15}$$

$$\widetilde{H}^{(N)} = \widetilde{A}^{(N)} \widetilde{H}^{(N+1)} \tag{B.16}$$

As a result, the transfer matrix relationship will connect the fields in the first layer to the fields in the last layer.

$$\widetilde{H}^{(1)} = \widetilde{A}^{(1)}\widetilde{A}^{(2)}\cdots\widetilde{A}^{(Last-1)}\widetilde{H}^{(Last)} \tag{B.17}$$

Up until now, this description is not significantly different than the transfer matrix method for dielectric stacks [1]. The only change that was made was that we assumed that the field was exponentially decaying or growing in each layer (that there was no propagation). In a purely dielectric stack, however, there is one extra degree of freedom. At the first interface, there is a forward propagating (incident) wave and a backward propagating (reflected) wave. In this plasmonic structure notation, a 'backward propagating' wave would be analogous to a term where the field grows exponentially away from the surface. If we assume that the first and last layers are unbounded, this will require that the fields are evanescently decaying away from the surface.

If region N = 1 is the left-most region, it should be unbounded and include negative infinity. Furthermore, there is one degree of freedom unaccounted for, so the N = 1 magnetic field can be normalized.

$$\widetilde{H}^{(1)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{B.18}$$

The last region should include positive infinity.

$$\widetilde{H}^{(Last)} = \begin{pmatrix} 0 \\ H_L \end{pmatrix} \tag{B.19}$$

If the system was symmetric, H_L would be restricted to +1 or -1. In an asymmetric structure, it will just be another unknown. Leaving only two unknowns, β and $H_{L,}$ with some finite number of possible solutions, $\beta^{(1)}$, $\beta^{(2)}$, etc. Thus the overall matrix equation can be written as follows.

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \widetilde{F}(\beta^{(i)}) \begin{pmatrix} 0 \\ H_L^{(i)} \end{pmatrix} \tag{B.20}$$

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} A(\beta^{(i)}) & B(\beta^{(i)}) \\ C(\beta^{(i)}) & D(\beta^{(i)}) \end{pmatrix} \begin{pmatrix} 0 \\ H_L^{(i)} \end{pmatrix} \tag{B.21}$$

This matrix equation suggests a rather simple solution.

$$D(\beta^{(i)}) = 0 \tag{B.22}$$

$$H_L^{(i)} = \frac{1}{B(\beta^{(i)})} \tag{B.23}$$

Because of the assumptions implicit in B.18 and B.19, this solution will provide only the bound states of a system.

REFERENCES

[1] A. Yariv, P. Yeh, *Photonics (6th Ed)*, New York: Oxford University Press, 2006.

[2] A.L. Fetter, J.D. Walecka, *Quantum Theory of Many-Particle Systems*, New York: Dover Publications, Inc., 2003.

[3] J.D. Jackson, *Classical Electrodynamics (3rd Ed)*, New York: John Wiley & Sons, Inc., 1999.

[4] H.A. Bethe, "Theory of Diffraction by Small Holes", The Physical Review 66, 163-182 (1944).

[5] P.B. Johnson, R.W. Christy, "Optical Constants of the Noble Metals", Phys. Rev. B 6, 4370-4379 (1972).

[6] E.D. Palik (Ed.), *Optical Constants of Solids*, New York: Academic Press, 1998.

[7] T.W. Ebbesen, H.J. Lezec, H.F. Ghaemi, T. Thio, P.A. Wolff, "Extraordinary Optical Transmission through Sub-Wavelength Hole Arrays", Nature 391, 667-669 (1998).

[8] Y. Takakura, "Optical Resonance in a Narrow Slit in a Thick Metallic Screen", Phys. Rev. Lett. 86, 5601-5603 (2001).

[9] M.M.J. Treacy, "Dynamical diffraction explanation of the anomalous transmission of light through metallic gratings", Phys. Rev. B 66, 195105(1)-(11) (2002).

[10] C. Genet, T.W. Ebbesen, "Light in Tiny Holes", Nature 445, 39-46 (2007).

[11] P. Anger, P. Bharadwaj, L. Novotny, "Enhancement and Quenching of Single-Molecule Fluorescence", Phys. Rev. Lett. 96, 113002(1)-(4) (2006).

[12] T. Okamoto, F. H'Dhilli, S. Kawata, "Towards plasmonic bandgap laser", Appl. Phys. Lett. 85, 3968-3970 (2004).

[13] Z. Sun, H.K. Kim, "Refractive transmission of light and beam shaping with metallic nano-optic lenses", Appl. Phys. Lett. 85, 642-644 (2004).

[14] H. Shi, C. Wang, C. Du, X. Luo, X. Dong, H. Gao, "Beam Manipulating by Metallic Nano-slits with Variant Widths", Optics Express 13, 6815-6820 (2005).

[15] J. Homola, S.S. Yee, G. Gauglitz, "Surface Plasmon resonance sensors: review", Sensors and Actuators B 54, 3-15 (1999).

[16] H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Berlin: Springer-Verlag, 1988.

[17] K.L. Tsakmakidis, C. Hermann, A. Klaedtke, C. Jamois, O. Hess, "Surface Plasmon Polaritons in Generalized Slab Heterostructures with Negative Permittivity and Permeability", Phys. Rev. B 73, 085104(1)-(11) (2006).

[18] K. Iizuka. *Elements of Photonics: Volume 2*, New York: John Wiley & Sons, Inc., 2002.

[19] L.A. Coldren, S.W. Corzine, *Diode Lasers and Photonic Integrated Circuits,* New York: John Wiley & Sons, Inc., 1995.

[20] P. Tournois, V. Laude, "Negative Group Velocities in Metal-Film Optical Waveguides", Opt. Comm. 137, 41-45 (1997).

[21] H. Shin, S. Fan, "All-Angle Negative Refraction for Surface Plasmons Using a Metal-Dielectric-Metal Structure", Phys. Rev. Lett. 96, 073907(1)-(4) (2006).

[22] I.V. Shadrivov, A.A. Sukhorukov, Y.S. Kivshar, "Guided Modes in Negative-Refractive-Index Waveguides", Phys. Rev. E 67, 057601(1)-(4) (2003).

[23] J. Wuenschell, H.K. Kim, "Surface Plasmon Dynamics in an Isolated Metallic Nanoslit", Opt. Express 14, 10000-10013 (2006).

[24] J. Nkoma, R. Loudon, D.R. Tilley, "Elementary Properties of Surface Polaritons", J. Phys. C: Solid State Phys. 7, 3547-3559 (1974).

[25] Z. Sun, Y.S. Jung, H.K. Kim, "Dynamic Evolution of Surface Plasmon Resonances in Metallic Nanoslit Arrays", Appl. Phys. Lett. 86, 023111(1)-(3) (2005).

[26] M. Scalora, M.J. Bloemer, A.S. Pethel, J.P. Dowling, C.M. Bowden, A.S. Manka, "Transparent, metallo-dielectric, one-dimensional, photonic band-gap structures", J. Appl. Phys. 83, 2377-2383 (1998).

[27] Y.S. Jung, Z. Sun, J. Wuenschell, H.K. Kim, P. Kaur, L. Wang, D. Waldeck, "High-Sensitivity Surface Plasmon Resonance Spectroscopy Based on a Metal Nanoslit Array", Appl. Phys. Lett. 88, 243105 (2006).

[28] D.M. Sullivan, *Electromagnetic Simulation Using the FDTD Method*, New York: IEEE Press, 2000.

[29] A. Taflove, S.C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method (3rd Ed.)*, Boston: Artech House, 2005.

[30] F.W. Byron, Jr., R.W. Fuller, *Mathematics of Classical and Quantum Physics*, New York: Dover Publications, Inc., 1970.

[31] E.N. Economou, "Surface Plasmons in Thin Films", Phys. Rev. 182, 539-554 (1969).

[32] J.A. Porto, F.J. Garcia-Vidal, J.B. Pendry, "Transmission Resonances on Metallic Gratings with Very Narrow Slits", Phys. Rev. Lett. 83, 2845-2848 (1999).

[33] S. Astilean, P. Lalanne, M. Palamaru, "Light Transmission through Metallic Channels Much Smaller than the Wavelength", Opt. Comm. 175, 265-273 (2000).

[34] P.N. Stayrinou, L. Solymar, "The Propagation of Electromagnetic Power through Sub-Wavelength Slits in a Metallic Grating", Opt. Comm. 206, 217-223 (2002).

[35] C. Liu, C. Yan, H. Chen, Y. Liu, S. Gao, "Evanescent Field on the Surface of a Negative-Index Planar Lens", Appl. Phys. Lett. 88, 231102(1)-(3) (2006).

[36] A.D. Boardman (Ed.), *Electromagnetic Surface Modes*, New York: John Wiley & Sons, Inc., 1982.

[37] H.F. Schouten, T.D. Visser, D. Lenstra, "Light transmission through a subwavelength slit: Waveguiding and optical vortices", Phys. Rev. E 67, 036608(1)-(4) (2003).